

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI INGEGNERIA E ARCHITETTURA

Corso di Laurea Magistrale in Ingegneria e Scienze Informatiche

**PREDIZIONE DEL TEMPO
RIMANENTE DI VITA DI UN
IMPIANTO MEDIANTE HIDDEN
MARKOV MODEL**

Relatore:

Chiar.mo Prof.

GOLFARELLI MATTEO

Presentata da:

Dott.sa

SAJEVA LISA

Sessione I Anno Accademico 2016-2017

Per comprendere e provare alcuni teoremi occorre estendere la propria mente oltre i suoi confini: l'informatica, la scienza dell'informazione e della computazione possono fare questo. E quando ci riescono, quello che si ottiene è di una bellezza astratta, quasi matematica nella sua perfezione: è come la musica, che è in grado di regalare a chiunque una esperienza assoluta".

Robin Milner

Indice

Introduzione	1
1 Condition-Based Maintenance	5
1.1 Reattivo, preventivo, predittivo	6
1.2 La diffusione	8
1.3 Condizioni	9
1.4 Approccio olistico	10
1.5 Gli obiettivi	11
1.6 Il bisogno da cui nasce la manutenzione predittiva	12
1.7 La manutenzione predittiva	13
1.8 Tassonomia degli approcci prognostici	15
1.8.1 Model-based	16
1.8.2 Data-driven	17
1.8.3 Reliability-based	17
1.9 Sintesi di vantaggi e svantaggi	18
2 Wavelets	23
2.1 Analisi dei segnali	23
2.2 Wavelets Analysis	24
2.3 Continuous Wavelets Transform	25
2.4 Discrete Wavelet Transform	26

2.5	Wavelet Packet Transform	29
2.6	Esempi di Wavelet Functions	30
2.6.1	Haar Wavelet	30
2.6.2	Meyer Wavelet	31
2.6.3	Morlet Wavelet	32
2.6.4	Daubechies Wavelet	32
3	Hidden Markov Model	35
3.1	Hidden Markov Model e fault diagnostic	36
3.2	Formulazione di un HMM	37
3.2.1	Forward algorithm	39
3.2.2	Viterbi algorithm	40
3.2.3	Baum-Welch algorithm	41
4	Metodo e Remaining Useful Life	43
4.1	Metodo basato su WPD e HMM	43
5	WPD e Fault Detection	51
5.1	Analisi dei dati	52
5.1.1	Comprensione di dati	52
5.1.2	Selezione dei dati	54
5.1.3	Preparazione ambiente di calcolo	54
5.1.4	Osservazioni preliminari	56
5.1.5	Interpretazione grafica	56
6	Applicazione di HMM	59
6.1	Ottimizzazione	62
6.2	Sequenza ottimale	63

7	Step per il calcolo di RUL	67
7.1	Verifica dell'accuratezza	67
7.2	Calcolo della RUL	71
7.3	Stima della bontà	73
8	Conclusioni	75
8.1	Confronto tra i modelli	75
8.2	Accurateza	76
9	Altri approcci al problema della manutenzione predittiva	77
9.1	Varianti	77
9.1.1	Diversa decomposizione dei dati	78
9.1.2	Diversa costruzione del modello	78
9.2	Alternative	78
9.2.1	PAC e LNPP	79
9.2.2	Cross-Entropy e Cross-Validation	79
9.2.3	LAD	80
9.2.4	SVM	80
A	Dati riassuntivi dei modelli a confronto	81
	Ringraziamenti	87
	Bibliografia	89

Introduzione

Nel mondo di oggi vengono prodotte e raccolte continuamente grandissime quantità di dati, non tutti sono rilevanti e non tutti sono impliciti.

Vogliamo qui concentrarci sui processi del mondo reale che in genere producono output osservabili che possono essere rappresentati come segnali. Questi segnali possono essere di natura discreta (come ad esempio i caratteri di un alfabeto finito, un certo numero di vettori di un codebook, ecc.) o continua (come ad esempio la temperatura, la musica, ecc.). Inoltre la sorgente del segnale può essere stazionaria (le proprietà del segnale non variano nel tempo) o non-stazionaria (le proprietà del segnale variano nel tempo). Il segnale può poi essere puro (se deriva esclusivamente da una singola sorgente) o corrotto da altre sorgenti (dette rumore) o da trasmissioni distorte, riverbero, ecc.. Per arginare questi problemi, possiamo contare su tecniche dette di *Data Mining* per costruire un modello a partire dai dati che permetta di ottenere le informazioni necessarie al controllo del suddetto processo. Col termine data mining ci si riferisce, in genere, all'intero processo, che si articola nei seguenti passi:

- la raccolta e analisi dei dati, dove l'analisi dei dati consiste nel capire come sono fatti i dati all'interno dei data set e decidere quali utilizzare al fine del data mining;
- lo sviluppo di modelli matematici di apprendimento che consentono di identificare se esiste una tendenza e permettono di rappresentarla sotto

forma di regola, di funzione o di grafico;

- l'adozione di decisioni e azioni basate sulle conoscenze acquisite.

In questa tesi sfrutteremo le tecniche di data mining per studiare la *Condition-Based Maintenance* (CBM) e nello specifico riprodurremo lo studio della *Remaining Useful Life* per un *Bearing System*. Il metodo prognostico usato in questo studio sarà di tipo data-driven, basato sull'impiego di gaussiane e Hidden Markov Model (MoG-HMM). L'utilizzo di questo strumento è motivato dal fatto che esso ci permette di gestire funzioni di densità di probabilità di emissione complesse (pdf) generate da un insieme di funzioni continue estratte da segnali di monitoraggio utilizzando Wavelet Packet decomposition (WPD). I dati a cui ci riferiremo sono presi dai data repository del *Prognostic Center of Excellence* della NASA[20]. Il modello verrà quindi costruito seguendo le seguenti fasi:

- analisi preliminare dei dati per comprendere il modo in cui sono stati registrati e il dominio applicativo a cui si riferiscono;
- trasformazione dei dati mediante *wavelet packet decomposition* per ottenere il segnale più significativo;
- costruzione del modello mediante Hidden Markov Model;
- uso del modello per determinare la Remaining Useful Life.

Nel Capitolo 1 saranno fornite alcune definizioni su CBM, failure prognostics e una tassonomia dei principali approcci nel campo della prognosi. Quest'ultima viene eseguita in due fasi: una fase off-line in cui i dati grezzi vengono utilizzati per imparare un modello comportamentale della condizione del componente fisica, e una fase on-line in cui viene utilizzato il modello appreso per identificare la condizione attuale del componente e per stimare la sua

RUL. Nel Capitolo 2 saranno introdotti i concetti teorici per l'elaborazione dei dati mentre nel Capitolo 3 quelli relativi al modello markoviano. Nei capitoli successivi il metodo è testato su dati di una simulazione reale relativi ai cuscinetti, e i risultati della simulazione sono riportati e discussi.

Capitolo 1

Condition-Based Maintenance

Finora le rivoluzioni industriali del mondo occidentale sono state tre: nel 1784 la macchina a vapore e il suo uso per meccanizzare la produzione; nel 1870 con la produzione di massa attraverso l'uso sempre più diffuso dell'elettricità e l'avvento del motore a scoppio; nel 1970 con la nascita dell'informatica, dalla quale è scaturita l'era digitale destinata ad incrementare i livelli di automazione avvalendosi di sistemi elettronici e dell'IT (Information Technology). La data d'inizio della quarta rivoluzione industriale non è ancora stabilita con certezza ma si può probabilmente attribuire all'anno 2016, anno in cui l'argomento al centro del *World Economic Forum* (Svizzera) era appunto "Mastering the Fourth Industrial Revolution". L'espressione Industria 4.0 è collegata appunto alla cosiddetta "quarta rivoluzione industriale". Tale rivoluzione è stata resa possibile dalla disponibilità di sensori e di connessioni wireless a basso costo, questa nuova rivoluzione industriale si associa a un impiego sempre più pervasivo di dati e informazioni, di tecnologie computazionali e di analisi dei dati, di nuovi materiali, componenti e sistemi totalmente digitalizzati e connessi (internet of things and machines). Secondo il *Ministero dello sviluppo economico* l'industria 4.0 richiede, per essere tale,

soluzioni tecnologiche per:

- ottimizzare i processi produttivi;
- supportare i processi di automazione industriale;
- favorire la collaborazione produttiva tra imprese attraverso tecniche avanzate di pianificazione distribuita, gestione integrata della logistica in rete e interoperabilità dei sistemi informativi.

In questo capitolo vogliamo concentrarci sugli aspetti di gestione dei guasti. Una delle possibili leve consiste nell'organizzare le attività di manutenzione; mantenendo il sistema in condizioni ottimali si possono ridurre i costi del ciclo di vita a livello globale, aumentare la sua disponibilità, migliorare la sicurezza degli operatori e ridurre gli incidenti ambientali.

La Condition Based Maintenance (CBM) è una filosofia di management che prende le decisioni su riparazione o sostituzione dei beni in base al loro stato presente o futuro. L'obiettivo è quello di minimizzare il costo totale delle ispezioni e delle riparazioni in base ai dati (continui o discontinui) relativi al funzionamento e alle condizioni delle componenti critiche di un'attività. Per farlo, un sistema di CBM deve fornire un adeguato preavviso in vista di un fallimento.

1.1 Reattivo, preventivo, predittivo

Le principali tecniche con cui possiamo prendere una decisione riguardo alla manutenzione si riferiscono al tipo di preavviso che si fornisce al fallimento e sono dette: manutenzione reattiva, manutenzione preventiva e manutenzione predittiva.

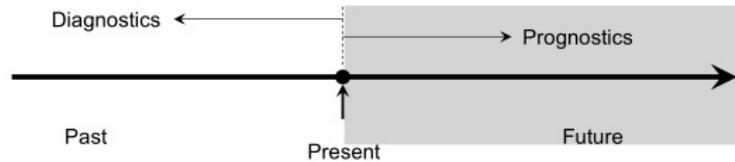


Figura 1.1: Diagnostico vs Prognostico.

La manutenzione reattiva tratta il caso in cui i componenti sono sostituiti solo quando non sono in grado di svolgere il compito per cui sono stati progettati. Il principale inconveniente di questa soluzione è che la macchina subisce il guasto, che a volte è semplicemente una situazione non desiderata (fuoriuscita di sostanze chimiche e avvelenamento, esplosione, ecc). Per superare queste situazioni, è possibile prevedere dei ricambi periodici, ovvero manutenzione preventiva, stimando la vita media del componente più fragile. Questo però a volte non è sufficiente, ad esempio nel caso di un guasto improvviso ci si può ritrovare nella situazione in cui i pezzi di ricambio non sono disponibili o non sono sufficienti, o semplicemente le risorse necessarie (manutentori) sono occupate. Inoltre questa tecnica ha lo svantaggio di non sfruttare appieno la vita di ciascun pezzo. Per ovviare a ciò si può pensare di monitorare alcuni parametri significativi del sistema, e quindi, modificando alcuni valori di soglia, si può procedere a sostituire i componenti quando i parametri monitorati superano le soglie definite, manutenzione predittiva. Questa azione fa parte del quadro di una manutenzione basata su condizioni (CBM). Ovviamente in termini di costo ed efficienza la manutenzione reattiva è nettamente peggiore delle precedenti, mentre tra le altre due la distinzione in questi termini non è così palese come vedremo (vedi tabella di vantaggi 1.1 e svantaggi 1.2).

1.2 La diffusione

Per decenni, specialmente nell'industria pesante, si è sentito parlare di manutenzione predittiva ma è solo grazie alla combinazione di sensori a basso costo, tecniche analitiche avanzate e IoT che questa tecnica è passata da costosa a proficua. Niente attira l'attenzione dei dirigenti come parlare di profitti, il che spiega il grande interesse corrente intorno appunto alla manutenzione predittiva. Le prime ad adottarla sono state le società nei settori dell'energia, trasporti, produzione, e tecnologia dell'informazione, infatti sono le società maggiormente dipendenti da servizi e prestazioni che scorrono attraverso la IoT. Uno studio dell'*U.S. Department of Energy's Pacific North-west National Laboratory* dichiara che i costi di manutenzione si riducono del 30%, quelli di inattività del 45% e i guasti fino al 75%.

Il CBM ha moltissime applicazioni nei campi più svariati. Le applicazioni industriali possono poi essere ulteriormente categorizzate a seconda del settore tra cui quello manifatturiero, automobilistico, dell'industria pesante, aerospaziale, della produzione di energia, dei trasporti, eccetera. In termini di successo, nel settore manifatturiero, ci sono alcuni risultati economici che possono essere citati. Ad esempio, la National Science Foundation ha finanziato uno studio da cui risulta che le 5 aziende (prevalentemente manifatturiere applicazioni) hanno riportato un risparmio di oltre 855 milioni di dollari grazie alla corretta attuazione del monitoraggio predittivo e delle soluzioni CBM [12].

Un altro esempio applicativo possono essere i veicoli pesanti utilizzati nella costruzione, nell'agricoltura e nell'estrazione mineraria. Alcuni dei maggiori produttori di queste apparecchiature come Komatsu e Caterpillar hanno già delle infrastrutture per il monitoraggio remoto, con algoritmi di analisi dei dati avanzati per rilevare i problemi del veicolo in una fase iniziale [13].

Gli esempi includono metodi e software per studio nel tempo dell'efficienza dei cuscinetti, per il monitoraggio del mandrino sulla base delle vibrazioni prodotte e del tipo di motore [11] in uso. In alcuni casi è messa a disposizione anche un'architettura cloud basata sul monitoraggio delle componenti come indicato sopra per mettere in relazione la salute con l'usura e la qualità dei pezzi. In genere vengono impiegati metodi basati su algoritmi di analisi e classificazione delle frequenze per rilevare i primi sintomi che vengono poi elaborati con tecniche di Data mining e approcci analitici avanzati.

I vantaggi della manutenzione predittiva sono quindi molti, permette di eliminare i fallimenti catastrofici dell'attrezzatura, di programmare le attività di manutenzione per minimizzare o eliminare i costi superflui, di ridurre al minimo l'inventario, di ordinare ricambi ben prima del tempo, di ottimizzare le operazioni delle apparecchiature risparmiando così costi energetici e migliorando l'affidabilità.

Compagnie come GE, Cisco, IBM e Intel contano sulle capacità di manutenzione predittiva fondate sull'IoT per creare altri \$ 100 miliardi di valore in energia e utilities entro il 2020. Secondo Gartner, l'IOT creerà quasi \$ 2 miliardi di dollari di nuovo valore in tutta l'economia globale nel corso del prossimo cinque anni e gran parte della creazione di valore sarà stimolato dalla manutenzione predittiva.

1.3 Condizioni

Uno dei requisiti fondamentali per poter applicare le tecniche di CBM è che il fallimento sia definito inequivocabilmente. Vogliamo qui considerare la seguente definizione di fallimento: *"il fallimento è la condizione di un bene che non è più pienamente in grado di adempiere alla sua missione"*.

Un altro punto chiave per la buona realizzazione di sistemi CBM è la bassa incidenza di falsi allarmi. Inoltre è molto importante anche il supporto manageriale a tutti i livelli (dati e analisi appropriate, personale addetto alla manutenzione competente).

Oltre a ciò, anche la scelta delle soglie influenza i costi di mantenimento, una soglia troppo bassa, che implica sostituire il pezzo più frequentemente, non permette di sfruttare al massimo la vita dei dispositivi mentre una troppo alta rischia di incorrere in fallimenti ovvero può capitare che il pezzo arrivi alla rottura prima che fosse prevista la sostituzione. Specialmente se l'ispezione è costosa, le date di ispezione e la soglia critica sono le principali variabili del problema di manutenzione. L'approccio conservativo può portare a soglie troppo deboli e ispezioni più frequenti del necessario, è quindi necessario determinare i valori ottimali.

Le tecniche di CBM potrebbero però non essere sempre convenienti in termini di costi-efficacia o potrebbero non esistere dati adeguati infatti, non esiste un'architettura generica per CBM ma ogni dominio ha una propria interpretazione, i CBM potrebbero non essere compatibili con i requisiti o con le altre applicazioni. É quindi sempre necessario svolgere prima un'analisi che permetta di stimare i vantaggi, i costi e l'applicabilità del sistema che si vuole adottare.

1.4 Approccio olistico

Un approccio possibile per l'implementazione di CBM è quello olistico, ovvero partendo dall'idea che le proprietà di un sistema non possono essere spiegate esclusivamente tramite le sue componenti.

Per metterlo in atto bisogna sviluppare degli indicatori di performance che costituiscano componenti e sistemi. La valutazione di questi si basa su:

1. Le performance fisiche,
2. La frequenza dei fallimenti,
3. Le attuali manutenzioni preventive messe in atto nei sistemi.

L'approccio olistico dovrebbe anche includere la misurazione delle prestazioni del sistema come strumento per costruire una strategia oggettiva come la valutazione di rottura (BD evaluation breackdown) dove prevale la filosofia della corsa al fallimento; la valutazione di PM scheduling dove pianificazione e programmazione sono rigorosamente applicate e la valutazione CBM dove la degradazione di beni critici e costosi è monitorata.

L'applicazione del CBM richiede l'utilizzo di alcuni strumenti analitici per le modalità di fallimento ed i suoi effetti e per l'analisi delle criticità (FMECA) e RCM, per determinare la probabilità di fallimento e il tipo di fallimento che si sarebbe verificato; nonché un affidabile strumento di informazione per catturare e monitorare le riparazioni ed i costi associati delle attività prese in considerazione.

FMECA è un processo sistematico per identificare tutti i modi possibili in cui i fallimenti di un elemento possono verificarsi e RCM è un approccio per identificare la più applicabile e conveniente manutenzione. RCM valuta ogni fallimento per determinarne le conseguenze e quindi per valutare il costo della manutenzione.

1.5 Gli obiettivi

Riassumendo, gli obiettivi principali di un sistema di manutenzione sono tre:

- ottimizzare la redditività aziendale;

1.6. IL BISOGNO DA CUI NASCE LA MANUTENZIONE PREDITTIVA

- incrementare l'efficienza abbassando in generale i costi delle operazioni;
- incrementare l'affidabilità e la disponibilità dei beni riducendo i costi di manutenzione.

Il primo obiettivo, ovvero quello di più alto livello, si concentra sul profitto e sull'impatto sul business. Il secondo ragiona in termini di ottimizzazione dell'intero funzionamento dell'impianto, non solo del bene fisico, quindi comprese le risorse umane. Per quanto riguarda il terzo punto invece, si tratta principalmente di ridurre a zero i tempi di inattività pianificati per ogni bene, il che consente un miglioramento di affidabilità e disponibilità.

Al livello più basso quindi ci si occupa di parti e macchinari, al livello successivo ci si pone tra risorse, macchine, processi e comportamenti umani e all'ultimo livello si verificano gli effettivi vantaggi e profitti portati dai primi due livelli.

1.6 Il bisogno da cui nasce la manutenzione predittiva

É possibile osservare che non sempre le revisioni basate su intervalli regolari di tempo aumentano l'affidabilità del sistema anzi a volte possono portare anche a disastri. Wood [17] come anche Moubray [18] hanno evidenziato che la manutenzione preventiva (PM) basata sul tempo spesso fallisce rispetto all'obiettivo di massimizzare la durata di ciascun componente, infatti in molti casi vengono sostituiti con ancora molte ore di vita residua (RUL). Questo significa che vengono sprecate molte risorse e che c'è sovra manutenzione.

È stato però possibile osservare che il 99% dei guasti è preceduto da segni o condizioni o da un qualche tipo di indicazione, perciò il monitoraggio può prevenire un guasto prima che diventi grave ed è quindi un'alternativa al PM basato sul tempo. Il CBM invece permette di focalizzarsi sulla sostituzione JIT (JustIn-Time) che massimizza la vita dei componenti.

La manutenzione predittiva è più di uno strumento o una soluzione, è una vera e propria strategia di business con molteplici strati, processi interconnessi e complesse relazioni.

La CBM non è una tecnica intrusiva inoltre può tenere in considerazione i fallimenti che spesso rimangono nascosti perché usati di rado e che si manifesterebbero solo al momento del bisogno, facendo uso di faultfiniding (FF) tasks che sono richiesti ad intervalli regolare per verificare lo stato di un bene ad esempio di un generatore di emergenza.

Un ostacolo al CBM è però il fatto che purtroppo molti managers non sanno quale sia il corretto regime di manutenzione e quindi molti di questi sistemi falliscono entro un anno spesso per la limitata comprensione del loro funzionamento e del dominio. Quest'ultima è fondamentale, infatti prendere decisioni in base agli allarmi ricevuti implica un'organizzazione per la manutenzione ben disciplinata.

1.7 La manutenzione predittiva

L'azione primaria che deve svolgere la manutenzione predittiva è quella di cercare di prevedere lo stato di salute del sistema e poi di prevedere azioni appropriate in base a ciò che annunciano le previsioni. Secondo l'International Standard Organization i *prognostici di fallimento* (o *failure prognostics*) sono la stima del tempo prima di un futuro fallimento e il rischio che esso avvenga. Esistono comunque molte altre definizioni in letteratura e tutte concordano sul

fatto che sia necessario stimare il tempo prima del guasto/fallimento. Questa stima viene indicata con la sigla RUL (Remaining Useful Life) in alcune opere mentre in altre come ETTF (Estimate Time To Failure) o ancora indicata come la probabilità che una macchina funzioni senza un guasto fino ad un certo momento futuro. In questa tesi adotteremo il termine RUL.

Diversamente dalla diagnostica dei guasti, che è un'attività matura, ben sviluppata e diffusa all'interno della ricerca e delle comunità industriali, la prognostica è un dominio di ricerca relativamente nuovo in particolare nella pratica.

Secondo quanto riportato in letteratura, i metodi di prognostica dei guasti possono essere classificati secondo tre principali approcci: model-based prognostics, reliability-based prognostics e data-driven prognostics.

I model-based prognostics si basa sull'utilizzo di un modello analitico, ovvero un insieme di equazioni algebriche o differenziali, per rappresentare il comportamento del sistema compreso il fenomeno di degradazione.

I reliability-based prognostics si basano sull'esperienza ovvero sui dati di feedback acquisiti durante un periodo di tempo significativo per regolare i parametri di alcuni modelli di affidabilità (Weibull, esponenziale, etc.). Questi ultimi sono quindi utilizzati per stimare il tempo di guasto, o RUL.

I data-based prognostics si concentrano sulla trasformazione, in modelli affidabili, dei dati forniti dai sensori in modo da catturarne sia il comportamento che la degradazione. Nella pratica questo metodo risulta particolarmente più facile e spesso anche meno complesso e più immediato nella comprensione rispetto a un modello analitico.

Oltre al valore assoluto della RUL, viene considerato anche un intervallo di confidenza, che è calcolato per tener conto dell'aspetto di incertezza insita nei pronostici di guasto. Infatti, esistono diversi fattori che possono influenza-

re il valore previsto del RUL, inoltre, in [19] è suggerito un elenco dei fattori e pesi corrispondenti, che possono influenzare il calcolo del valore di confidenza. La Figura 1.2 illustra il RUL, e il grado di fiducia associato, in questa figura si nota che l'incertezza può essere di due tipi: il primo è dovuto alla previsione, e il secondo è legato al valore di soglia corrispondente al fallimento completo della macchina.

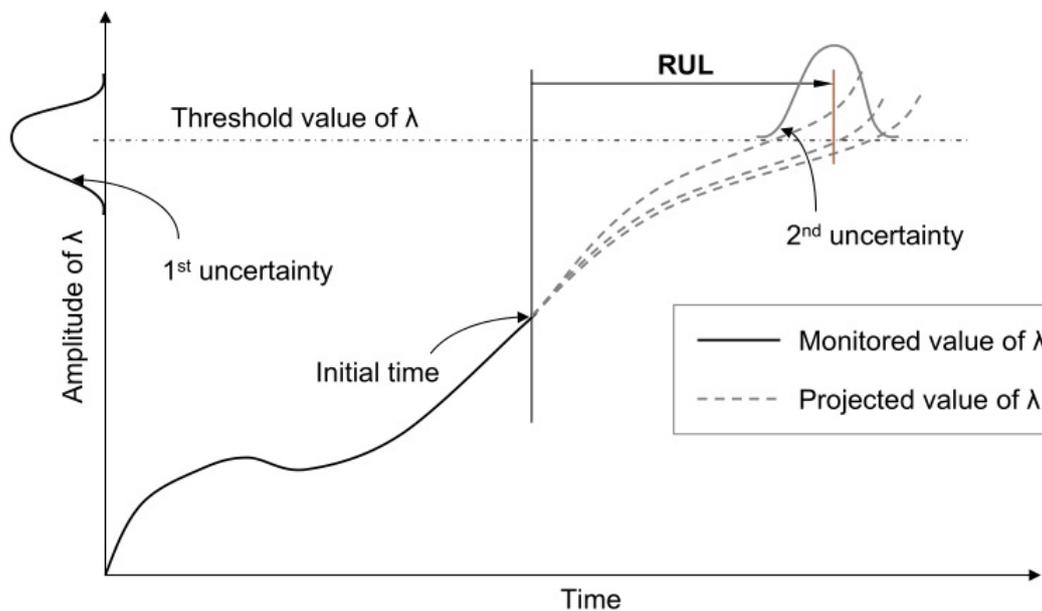


Figura 1.2: Incertezza relativa a RUL.

1.8 Tassonomia degli approcci prognostici

Il fallimento ed in particolare la stima della vita utile residua, può essere fatto utilizzando numerosi strumenti e metodi. Questi ultimi possono essere raggruppati in tre approcci principali come accennato prima (vedi Fig. 1.3),

Questa classificazione è basata sul tipo di dati e sugli strumenti coinvolti, inoltre, come si può notare, possono esistere intersezioni tra i tre approcci.

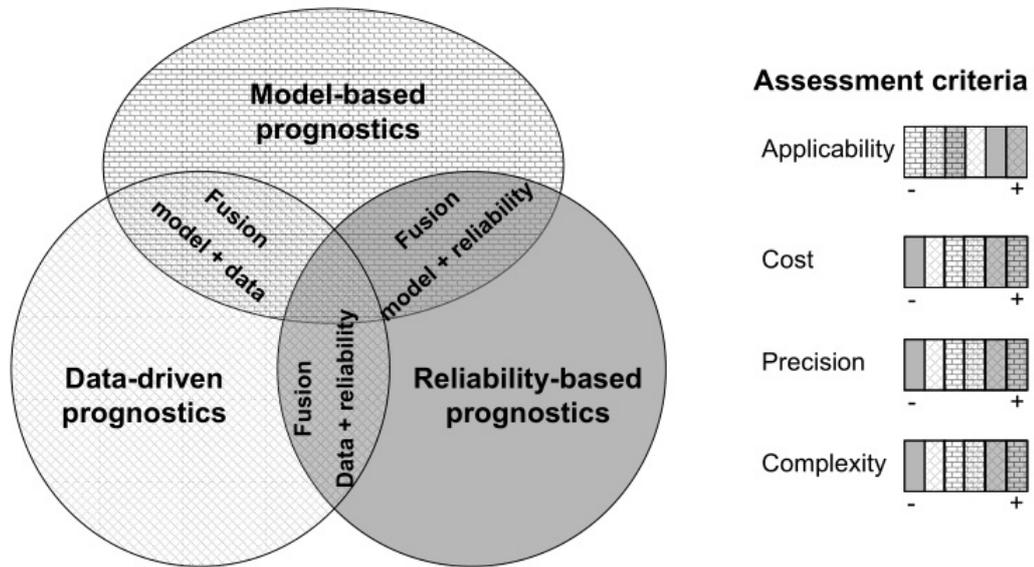


Figura 1.3: Principali approcci prognostici.

Allo stesso modo è possibile utilizzare più di uno strumento a seconda del dominio di applicazione. Vedremo ora una breve descrizione di ogni approccio seguito da una sintesi dei vantaggi e degli inconvenienti legati a ciascuno di essi.

1.8.1 Model-based

Come accennato prima questo approccio si basa sull'uso di un modello analitico che descriva il comportamento della componente in esame. Tale modello analitico può essere un insieme di equazioni algebriche e/o differenziali ottenuti utilizzando tradizionali leggi della fisica. Il fenomeno di degradazione è rappresentato da una o più variabili la cui dinamica è regolata da un insieme di parametri dipendenti dal contesto in cui il sistema fisico si evolve. In alcune applicazioni (modelli di stato per esempio), la variabile che rappresenta

la degradazione può essere considerata parte del modello di comportamento globale.

Si noti che, in pratica, i modelli di propagazione delle fratture/guasti presuppongono che sia possibile determinare la dimensione della frattura mediante opportuni sensori (analisi vibrazioni per esempio) perché una verifica visiva è spesso difficile, se non impossibile.

1.8.2 Data-driven

I metodi di questo approccio mirano a trasformare i dati di monitoraggio non elaborati in informazione e quindi in modelli di comportamento. Il modello di degradazione è derivato utilizzando i soli dati forniti dal sistema di monitoraggio, formato principalmente da sensori, senza preoccuparsi del modello analitico del sistema né dei suoi parametri fisici (come proprietà del materiale). I metodi di prognostica data-driven utilizzano principalmente gli strumenti dell'intelligenza artificiale (reti neurali, reti bayesiane, processi Markoviani, etc.) o i metodi statistici (modelli di regressione, reti Bayesiane dinamiche attraverso filtri di Kalman e particelle, etc.). Per conoscere il modello di degrado, e per prevedere il futuro stato di salute del sistema.

Il principio di questi metodi si compone di due fasi: una prima fase nella quale viene appreso un modello di comportamento (compresa la degradazione) in genere sviluppata off-line; e una seconda fase in cui viene utilizzato il modello appreso per stimare dapprima la condizione di funzionamento attuale del sistema, e quindi per prevedere il suo futuro stato in genere on-line.

1.8.3 Reliability-based

I metodi di prognostica basati sull'esperienza utilizzano principalmente i dati di feedback dell'esperienza acquisita durante un periodo di tempo significa-

tivo (dati di manutenzione, dati operativi, tempi di guasto, ecc) per regolare i parametri di alcuni modelli di affidabilità predefiniti. I modelli ottenuti sono poi utilizzati per prevedere il tempo mancante al fallimento. In letteratura sono stati proposti diversi modelli parametrici di fallimenti: Poisson, esponenziale, Weibull, e le leggi di log-normale. Tra questi modelli, la distribuzione di Weibull è la più segnalata, in quanto può rappresentare diverse fasi temporali della vita del componente o sistema

1.9 Sintesi di vantaggi e svantaggi

I tre principali approcci alla prognostica verranno confrontati qui di seguito secondo quattro criteri: precisione, complessità, costi e applicabilità.

Rispetto agli approcci basati sull'esperienza, quelli basati sui dati e sui modelli danno risultati di prognostica più precisi. Inoltre la stima del valore di confidenza associato alla RUL è facile da ottenere cambiando i parametri fisici dei modelli di degrado, e l'utilizzo di metodi di simulazione statistici come Monte-Carlo. I metodi basati su modelli sono facili da interpretare perché i parametri nei modelli comportamentali corrispondono ai fenomeni fisici che avvengono nei sistemi. Tuttavia, nonostante la precisione delle prognosi che questi metodi forniscono, è in pratica difficile generare un modello per il comportamento del degrado, specialmente per sistemi complessi dove avvengono diversi tipi di fenomeni fisici. Infatti, anche se il modello esiste, è generalmente una specifica rappresentazione di un fenomeno fisico generato in condizioni specifiche e sperimentazioni. Così, ripetere gli esperimenti per le diverse condizioni di funzionamento può essere costoso, il che limita l'applicabilità di questo approccio.

L'approccio basato sull'esperienza è facile da applicare ai sistemi in cui sono disponibili i dati rappresentativi di utilizzo raccolti durante un lungo

periodo di tempo. Il compito principale di questo approccio consiste infatti nella stima dei parametri delle leggi di affidabilità noti dai dati storici. I metodi appartenenti a questo approccio sono anche meno costosi da implementare tuttavia i risultati di prognostica sono meno precisi di quelli ottenuti con metodi basati sui modelli e sui dati. Pertanto l'utilizzo di metodi basati sull'esperienza non è auspicabile per i sistemi in cui i risultati prognostici sono critici. Inoltre, questo approccio è di difficile applicazione nel caso dei nuovi sistemi a causa della mancanza di dati di esperienza.

Lo sviluppo di sensori e informatica ha facilitato l'uso di intelligenza artificiale (AI), e di conseguenza l'uso di metodi basati sui dati. Questi metodi trasformano i dati di monitoraggio in modelli comportamentali utilizzando appunto strumenti di intelligenza artificiale. L'approccio basato sui dati offre un compromesso in termini di complessità, costo, precisione e applicabilità. Infatti, rispetto agli approcci basati su modelli, sono adatti a sistemi in cui è facile ottenere dati di monitoraggio che rappresentano il comportamento del fenomeno di degradazione. Nella pratica, questa condizione è frequente in diverse applicazioni, come i cuscinetti, che sono oggetto del presente documento. È vero che esistono modelli di previsione, come ad esempio la legge L10, per calcolare la vita utile residua dei cuscinetti, ma questa legge è valida solo per cuscinetti in specifiche condizioni che sono difficili da verificare nelle applicazioni del mondo reale. Così, per superare questa situazione, i metodi basati sui dati possono essere considerati come una soluzione alternativa. Lo svantaggio dei metodi basati sui dati, almeno per alcune applicazioni, è il tempo di apprendimento potenzialmente lungo. In termini di precisione, i metodi basati sui dati danno risultati meno precisi rispetto a quelli basati sui modelli, ma meglio di quelli basati sull'esperienza, tuttavia, i metodi basati sui dati sono meno complessi e più applicabili di quelli basati sui modelli. Di

conseguenza, nel seguito, si propone un metodo basato sui dati per prevedere RUL dei cuscinetti.

Tabella 1.1: Vantaggi della manutenzione preventiva e predittiva ([16]).

Manutenzione Preventiva	Manutenzione Predittiva
Costo effettivo misurabile in molti processi ad alta intensità di capitale	Maggiore disponibilità/durata di vita operativa del componente
Flessibilità che consente di regolare la periodicità della manutenzione	Permette azioni di correzione preventiva
Aumento del ciclo di vita delle componenti	Calo dei costi per le parti di ricambio e la mano d'opera
Risparmio energetico	Risparmio energetico
Si stima un risparmio tra il 12% e il 18% rispetto alla manutenzione programmata	Si stima un risparmio tra l'8% e il 12% rispetto al programma di manutenzione preventiva
Riduce l'equipaggiamento o il guasto del processo	Prodotti di migliore qualità
	Minor equipaggiamento necessario e minor tempo di downtime
	Miglioramento del lavoro e della sicurezza ambientale
	Migliora il morale dei lavoratori

Tabella 1.2: Svantaggi della manutenzione preventiva e predittiva ([16]).

Manutenzione Preventiva	Manutenzione Predittiva
I fallimenti catastrofici potrebbero ancora verificarsi	Aumento degli investimenti in attrezzatura diagnostica
lavoro intensivo	Crescita degli investimenti nella formazione
Presenza di manutenzioni non necessarie	Risparmio potenziale non facilmente individuabile dal management
Rischio potenziale di danni ai componenti durante manutenzioni non necessarie	

1.9. SINTESI DI VANTAGGI E SVANTAGGI

Capitolo 2

Wavelets

Sebbene la teoria delle wavelets non sia nuova, i recenti risultati permettono di trattare diversi problemi di interesse in diversi campi quali: l'ingegneria elettrica (analisi dei segnali, compressione dei dati), l'analisi matematica e la fisica [1]. Vogliamo qui occuparci dell'analisi dei segnali, specialmente di quelli non stazionari che negli ultimi anni ha suscitato grande interesse per via della loro grande presenza in situazioni reali.

2.1 Analisi dei segnali

I segnali non stazionari sono caratterizzati da andamenti di lungo periodo a basse frequenze detti anche *trend*, e andamenti di breve periodo ad alte frequenze detti anche *anomalies*. Queste ultime, sebbene occupino una percentuale relativamente piccola dei dati, presentano un elevato contenuto informativo, e vanno quindi rappresentati adeguatamente. Tale rappresentazione può essere ottenuta in modo accurato sfruttando appunto il meccanismo delle Wavelets. La trasformata wavelet è un potente strumento per l'analisi e l'elaborazione dei segnali e risulta estremamente efficiente in diversi campi di

applicazione, come la compressione e il denoising e in generale quando si ha a che fare con segnali non stazionari, come sono ad esempio le immagini o le vibrazioni. Possiamo immaginare il risultato di una wavelet transform come un grafico 3D che abbia per assi frequenza, ampiezza e tempo. Ragionando sulle caratteristiche di un segnale possiamo osservare che le frequenze alte sono meglio risolvibili in termini di tempo, infatti frequenze alte per lo stesso intervallo di tempo implicano campioni più numerosi, mentre quelle basse sono meglio descritte in termini di ampiezza. Questo fatto è riassunto dal principio di indeterminazione di Heisenberg e poiché l'oggetto della Wavelet Transform è un'onda, anch'essa è soggetta a tale principio. Questo ci fa capire come si adattino perfettamente ai segnali di tipo non stazionario di nostro interesse in questa trattazione.

2.2 Wavelets Analysis

Come l'analisi di Fourier anche l'analisi wavelets sfrutta la decomposizione di funzioni in termini di un insieme di funzione di base ma invece di utilizzare polinomi trigonometrici (sinusoidi complesse), la wavelet transform decompone la funzione in una famiglia di wavelets che a differenza delle sinusoidi che sono simmetriche, smooth e regolari, possono essere sia simmetriche che asimmetriche, sia sharp che smooth, sia regolari che irregolari. A seconda del tipo di funzione di ingresso quindi, si sceglierà un diverso tipo di wavelets, quello che meglio si adatta alle caratteristiche della funzione prototipo detta *mother wavelet*. Qualsiasi famiglia di wavelet si scelga conterrà quindi la versione dilatata e traslata della funzione prototipo. La dilatazione e la traslazione si riferiscono rispettivamente all'asse del tempo e dello spazio.

La Wavelet Transform presentano quindi interessanti proprietà:

- Sono in grado di analizzare segnali con caratteristiche variabili nel tempo e nella frequenza.
- Forniscono una rappresentazione su diverse scale.
- Sono finite nel tempo.
- Possono analizzare sia dati monodimensionali che multidimensionali.
- Il processo è reversibile.

Si possono distinguere due macro categorie di Wavelet Transform ovvero continue o discrete. Di solito, si utilizzano gli strumenti wavelet continui per l'analisi dei segnali, come l'analisi di auto-similarità e l'analisi tempo-frequenza e gli strumenti wavelet discreti sia per l'analisi dei segnali che per l'elaborazione del segnale, come ad esempio la riduzione del rumore, la compressione dei dati, il rilevamento di picchi e così via.

2.3 Continuous Wavelets Transform

Sia ψ una funzione a valori complessi che soddisfi le seguenti proprietà:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \quad (2.1)$$

$$c_{\psi} = 2\pi \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad (2.2)$$

dove Ψ è la trasformata di Fourier di ψ . La prima condizione 2.1 indica che la funzione ψ ha energia finita mentre la seconda 2.2 è una condizione di ammissibilità ed implica che se $\Psi(\omega)$ è smooth allora $\Psi(0) = 0$.

La funzione appena descritta è la *mother wavelet* e la *Wavelet Transform* di un segnale reale $s(t)$ rispetto ad $\psi(t)$ è definito come:

$$S(b, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi' \left(\frac{t-b}{a} \right) s(t) dt \quad (2.3)$$

dove ψ' è il complesso coniugato di ψ ed è definito sul semipiano aperto (b, a) con $b \in \mathbb{R}, a > 0$. Nella nostra notazione il parametro b corrisponde al variare del tempo ed a al variare della scala della wavelet analizzata.

Chiamiamo ora $\psi_{a,b}(t)$ il seguente valore:

$$\psi_{a,b}(t) = a^{-\frac{1}{2}} \psi \left(\frac{t-b}{a} \right) \quad (2.4)$$

che significa *scaling* rispetto ad a e *shifting* rispetto a b allora possiamo riscrivere l'equazione 2.3 come il prodotto del segnale reale $s(t)$ e della funzione $\psi_{a,b}(t)$:

$$S(b, a) = \int_{-\infty}^{\infty} \psi'_{a,b}(t) s(t) dt \quad (2.5)$$

Quando la funzione $\psi(t)$ soddisfa la condizione di ammissibilità 2.2, il segnale originale $s(t)$ può riottenere tramite la funzione inversa della Wavelet Transform $S(b, a)$ con la seguente formula:

$$s(t) = \frac{1}{c_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(b, a) \psi_{a,b}(t) \frac{dad b}{a^2} \quad (2.6)$$

2.4 Discrete Wavelet Transform

Se nelle Continuous Wavelet Transform (CWT) il segnale viene analizzato usando un insieme di funzioni di base correlate tra loro da semplici scalature o traslazioni, nel caso delle Discrete Wavelets Transform (DWT) una rappresentazione scalare del tempo per il segnale digitale si ottiene usando tecniche di filtri digitali. Il segnale per essere analizzato passa attraverso filtri con differenti tagli di frequenze e scale diverse. Quando un segnale passa attraverso

2. WAVELETS

un filtro esso si divide in due bande (vedi Figura 2.4). Questo processo ha due vantaggi, aumenta la frequenza della risoluzione e dimezza l'incertezza sulla frequenza.

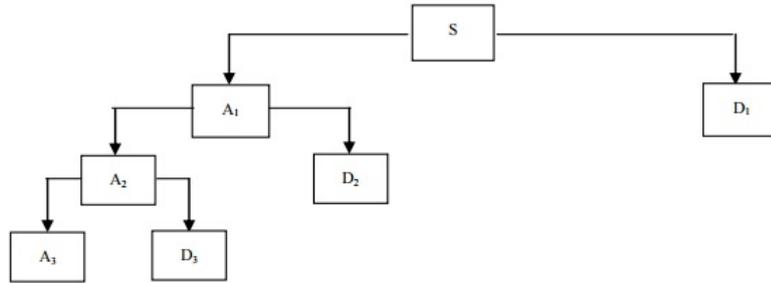


Figura 2.1: Decomposizione usando wavelet transform a 3 livelli.

L'idea alla base è quindi quella di suddividere iterativamente il segnale mediante filtri passa alto e passa basso. Tali filtri sono descritti come segue.

Siano

$h(n)$ un low – pass filter (L) relativo alla funzione scalare $\Phi(t)$

e

$g(n)$ un high – pass filter (H) relativo alla funzione wavelet $\Psi(t)$

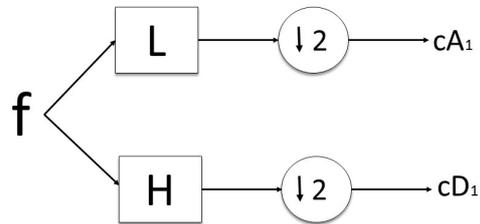
dove

$$\Phi_j(t) = \sum_k h(k) 2^{\frac{j+1}{2}} \Phi(2^{j+1}t - k) \quad (2.7)$$

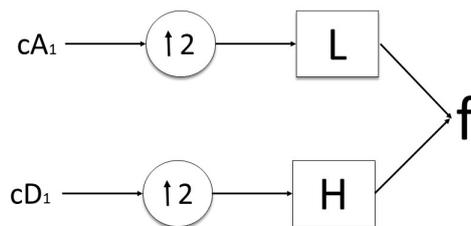
e

$$\Psi_j(t) = \sum_k h(k) 2^{\frac{j+1}{2}} \Psi(2^{j+1}t - k) \quad (2.8)$$

Allora il segnale f viene filtrato da H e L come in Figura 2.2 e si ottengono i vettori cA_1 e cD_1 che rappresentano rispettivamente l'*approximation coefficient* e il *detail coefficient*.



(a) Campionamento verso il basso



(b) Campionamento verso l'alto

Figura 2.2: Schema della decomposizione e ricomposizione del segnale.

Per passare al dominio discreto sia i parametri che le wavelets della formula 2.4. Assumiamo i seguenti cambi di variabili:

$$a = a_0^m \tag{2.9}$$

$$b = nb_0 \tag{2.10}$$

Allora la formula 2.4 diventa:

$$\psi_{m,n}(t) = a_0^{-\frac{m}{2}} \psi\left(\frac{t - nb_0}{a_0^m}\right) \quad (2.11)$$

con $m, n \in \mathbb{N}$. Analogamente al caso continuo anche per il caso discreto abbiamo a disposizione sia Discrete Wavelet Transform (DWT) e la sua inversa. Ed è quindi possibile esprimere il segnale come segue:

$$F_{m,n} = \sum_m \sum_n \psi'_{m,n}(t) f(t) \quad (2.12)$$

$$f(t) = k_\psi \sum_m \sum_n F_{m,n} \psi_{m,n}(t) \quad (2.13)$$

dove k_ψ è un valore costante per la normalizzazione.

La funzione $\psi_{m,n}(t)$ fornisce i punti di campionamento nel piano scala-tempo: campionamento lineare nel tempo (asse b) e logaritmico nella scala (asse a).

La scelta per a_0 più comune è quella di prendere

$$a_0 = 2^{\frac{1}{v}} \quad (2.14)$$

dove $v \in \mathbb{N}$.

2.5 Wavelet Packet Transform

Una Wavelet Packet Transform (WPT) è una generalizzazione della Wavelet Transform che offre una vasta scelta di possibilità per l'analisi dei segnali e che si adatta meglio ad essa. Infatti ad ogni passo entrambi i filtri vengono applicati (vedi Figura 2.3), in questo modo è possibile studiare anche le alte frequenze con miglior risoluzione infatti il segnale risulta diviso in coefficienti di *approximation* e *detail*.

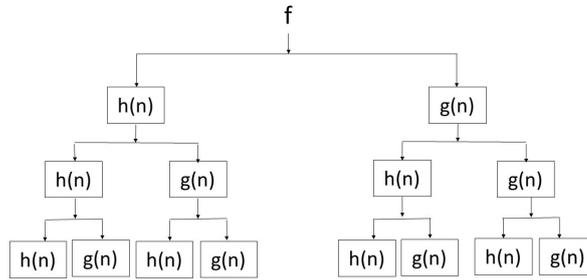


Figura 2.3: Schema dell'albero costruito dalla wavelet transform

Questo processo porta a più di 2^{2^n-1} differenti modi di codificare il segnale [2]. La capacità di iterare entrambi i tipi di filtri implica che la WPT permette di usare più di una funzione di base per ogni data scala. Il top level dell'albero di WPD è la rappresentazione del tempo per il segnale, mentre il bottom level di un albero completo è la rappresentazione della frequenza del segnale.

Se la complessità computazionale della Wavelet Transform è $O(N)$ quella della WPT è equivalente a quella della Trasformata di Fourier ovvero a $O(N \log_2 N)$

2.6 Esempi di Wavelet Functions

In questa sezione vedremo quattro dei più comuni tipi di wavelet function che variano poi comunque in funzione dei parametri di dilatazione scelti: Haar Wavelet, Mayer Wavelet e Daubechies Wavelet.

2.6.1 Haar Wavelet

Le Haar wavelets sono le funzioni wavelet più semplici, le prime ad essere studiate e le uniche finché l'interesse sulle wavelet non esplose. Supponiamo

che $\phi(t)$ sia una box function tale che

$$\phi(t) = \begin{cases} 1 & \text{se } 0 \leq t \leq 1 \\ 0 & \text{altrimenti} \end{cases} \quad (2.15)$$

essa è detta *funzione scalare di Haar*, definiamo quindi

$$\psi(t) = \phi(2t) - \phi(2t - 1) \quad (2.16)$$

e otteniamo la *funzione wavelet di Haar* come

$$\psi(t) = \begin{cases} 1 & \text{se } 0 < t \leq 1/2 \\ -1 & \text{se } 1/2 < t \leq 1 \\ 0 & \text{altrimenti} \end{cases} \quad (2.17)$$

Tali funzioni sono ortogonali rispetto alla loro traslazione e dilatazione.

2.6.2 Meyer Wavelet

Yves Meyer costruì la wavelet che porta il suo nome in modo che fosse ortonormale e smooth. Il primo passo per ottenerla è definire la trasformata di Fourier $\Phi(\omega)$ di una funzione scalare $\phi(t)$ come segue:

$$\Phi(\omega) = \begin{cases} 1 & \text{se } |\omega| \leq \frac{2}{3}\pi \\ \cos \left[\frac{\pi}{2} v \left(\frac{3}{4\pi} |\omega| - 1 \right) \right] & \text{se } \frac{2}{3}\pi < |\omega| \leq \frac{4}{3}\pi \\ 0 & \text{altrimenti} \end{cases} \quad (2.18)$$

dove v è una funzione smooth che soddisfa le seguenti condizioni:

$$v(t) = \begin{cases} 0 & \text{se } t \leq 0 \\ 1 & \text{se } t \geq 0 \end{cases} \quad (2.19)$$

e la proprietà

$$v(t) + v(1-t) = 1 \quad (2.20)$$

2.6.3 Morlet Wavelet

Questa particolare funzione fu la più usata da R. Kronland-Martinet e J. Morlet nei loro studi. La sua trasformata di Fourier è una Gaussiana traslata ed aggiustata in modo che $\Psi(\omega) = 0$:

$$\Psi(\omega) = e^{-(\omega-\omega_0)^2/2} - e^{-\omega^2/2} e^{-\omega_0^2/2} \quad (2.21)$$

$$\psi(t) = (e^{-i\omega_0 t} - e^{-\omega^2/2}) e^{-t^2/2} \quad (2.22)$$

2.6.4 Daubechies Wavelet

Ad eccezione di Haar, tutte gli esempi di wavelet ortonormali hanno funzioni a supporto infinito. Ingrid Daubechies costruì invece una wavelet ortonormale con ϕ a supporto compatto. Per farlo ha ovviamente scelto una funzione scalare Φ a supporto compatto.

Prima di tutto si trova una progressione $\{a_k | k \in \mathbb{Z}\}$ che soddisfi le seguenti quattro condizioni per tutti gli interi $N \geq 2$

$$a_k = 0 \text{ se } k < 0 \text{ o } k > 2N \quad (2.23)$$

$$\sum_{k=-\infty}^{\infty} a_k a_{k+2m} = \delta_{0m} \forall m \in \mathbb{N} \quad (2.24)$$

$$\sum_{k=-\infty}^{\infty} a_k = \sqrt{2} \quad (2.25)$$

2. WAVELETS

$$\sum_{k=-\infty}^{\infty} \beta^k k^m = 0 \text{ se } 0 \leq m \leq N-1 \quad (2.26)$$

dove $\beta_k = (-1)^k \alpha_{-k+1}$.

Il caso con $N = 1$ e $\alpha_0 = \alpha_1 = 1$ corrisponde alla Haar function. Si può quindi calcolare $\phi(t)$ a supporto compatto dalla progressione a_k come:

$$\phi(t) = \sum_{k=-\infty}^{\infty} a_k \sqrt{2} \phi(2t - k) \quad (2.27)$$

Capitolo 3

Hidden Markov Model

La teoria delle catene di Markov è nata a metà del XX secolo. Si definisce *processo stocastico markoviano* (o di Markov), un processo aleatorio nel quale la probabilità di transizione da uno stato all'altro del sistema dipende solo dallo stato immediatamente precedente e non dal dall'intero percorso. Possiamo quindi definire una *catena di Markov* come un processo di Markov che assume valori in uno spazio discreto che può essere sia finito che infinito. A questo punto possiamo immaginare un Hidden Markov Model (HMM) come una catena di Markov in cui gli stati non sono osservabili direttamente ovvero avremmo la seguente situazione:

- la catena ha un certo numero di stati
- gli stati evolvono secondo una catena di Markov
- ogni stato genera un evento con una certa distribuzione di probabilità che dipende solo dallo stato
- l'evento è osservabile ma lo stato no

Gli HMM sono diventati popolari di recente per le loro applicazioni nel riconoscimento dello schema temporale dei discorsi parlati, della scrittura a

mano, nel riconoscimento di texture e nella bioinformatica. Questo per due particolari ragioni, da una parte perché richiedono una struttura matematica molto ricca e dall'altra i modelli hanno molte applicazioni pratiche anche grazie alla semplicità della loro interpretazione grafica.

3.1 Hidden Markov Model e fault diagnostic

Il meccanismo di fallimento dei sistemi meccanici di solito coinvolge diversi stati di degrado della salute del sistema. Un tale processo può essere descritto con un HMM che può essere usato per stimare gli stati di salute non osservabili tramite segnali osservabili derivanti da sensori. In altre parole gli stati dell'HMM si manifestano grazie a comportamenti probabilistici, si può quindi rappresentare le transizioni tra gli stati come in Figura 3.1 dove le frecce rappresentano le transizioni tra uno stato e l'altro ed in genere a ciascuna è associato un numero che denota la probabilità associata alla transizione.

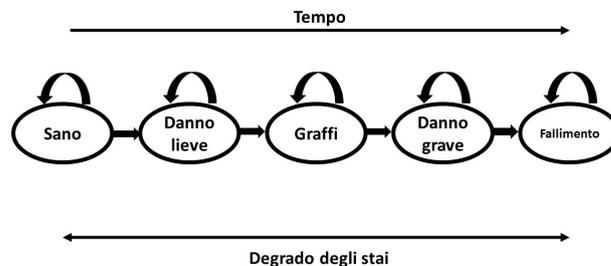


Figura 3.1: Schema per un HMM che descrive il meccanismo di fallimento di un processo meccanico.

Il segnale oggetto dell'analisi deve essere preprocessato per individuarne le caratteristiche relative allo specifico fallimento. Allo scopo di individuare la presenza di guasti, è sufficiente addestrare un singolo HMM per le condi-

zioni normali che possono essere prese da dati a partire da diverse condizioni iniziali. Data poi la matrice con le caratteristiche dei guasti vengono calcolate le probabilità dell'HMM. Se tali probabilità sono sopra una certa soglia allora non c'è fallimento e viceversa.

3.2 Formulazione di un HMM

Ogni HMM è definito da una quintupla di elementi: l'insieme degli stati, le probabilità degli stati, le probabilità di transizione, la probabilità finale ed iniziale.

1. Sia S l'insieme contenente gli N stati del modello

$$S = \{S_1, \dots, S_N\} \quad (3.1)$$

2. Sia V l'insieme degli M simboli per le osservazioni

$$V = \{v_1, \dots, v_M\} \quad (3.2)$$

Se le osservazioni sono continue M è infinito.

3. Sia $A = \{a_{ij}\}$ la distribuzione di probabilità delle transizioni tra stati, dove a_{ij} è la probabilità che lo stato al tempo $t + 1$ sia S_j se lo stato al tempo t è S_i . La struttura di questa matrice stocastica definisce la struttura delle connessioni del modello. Se un coefficiente a_{ij} è zero rimarrà zero per l'intero processo ovvero non ci saranno transizioni da S_i a S_j infatti

$$p(q_{t+1} = j | q_t = i) = S_j \cdot a_{ij} \quad (3.3)$$

Le transizioni devono quindi soddisfare le seguenti proprietà:

- (a) $a_{ij} \geq 0$ con $1 \leq i, j \leq N$

$$(b) \sum_{j=1}^N a_{ij} = 1 \text{ con } 1 \leq i \leq N$$

4. Sia $B = \{b_j(k)\}$ la distribuzione di probabilità dei simboli delle osservazioni in ogni stato dove $b_j(k)$ è la probabilità che il simbolo v_k sia sviluppato nello stato S_j

$$b_j(k) = p(o_t = v_k | q_t = j), \quad 1 \leq j \leq N, \quad 1 \leq k \leq M \quad (3.4)$$

dove v_k denota il k -esimo simbolo di osservazione nell'alfabeto e o_t il vettore corrente di parametri. I seguenti vincoli stocastici devono essere soddisfatti:

$$(a) b_j(k) \geq 0 \text{ con } 1 \leq j \leq N, \quad 1 \leq k \leq M$$

$$(b) \sum_{k=1}^M b_j(k) = 1 \text{ con } 1 \leq j \leq N$$

In genere la densità di probabilità viene approssimata pesando la somma di M distribuzioni Gaussiane $N(\mu, \sigma)$.

5. Sia $\pi = \{\pi_i\}$ dove π_i è la probabilità che il modello sia nello stato S_i al tempo $t = 0$ con

$$\pi_i = p(q_1 = i) \text{ con } 1 \leq i \leq N \quad (3.5)$$

Oltre alla definizione appena descritta ne esiste anche una più compatta ovvero:

$$\lambda = (A, B, \pi) \quad (3.6)$$

La modellazione stocastica di un automa HMM consiste di due passi:

1. Definire l'architettura del modello, ovvero estrarre un insieme di caratteristiche dai dati grezzi (vedi [10]).
2. Definire gli algoritmi di learning e operating, ovvero assegnare una classe agli stati dell'HMM.

Per quanto riguarda il punto 2 esistono 3 problemi canonici connessi con gli HMM:

1. **Evaluation Problem** Dati i parametri del modello in input, calcolare la probabilità che una particolare sequenza di osservazioni $O = o_1, o_2, \dots, o_T$ sia generata dal modello ovvero $p(O|\lambda)$. (algoritmo forward-backward)
2. **Decoding Problem** Dato il modello λ e una sequenza di osservazioni $O = o_1, o_2, \dots, o_T$ in input, calcolare la sequenza più probabile di stati che produce l'osservazione data. (algoritmo di Viterbi)
3. **Learning Problem** Dato il modello λ e una sequenza di osservazioni $O = o_1, o_2, \dots, o_T$ in input, aggiustare i parametri $\{A, B, \pi\}$ del modello in modo da massimizzare $p(O|\lambda)$ (algoritmo di Baum-Welch).

3.2.1 Forward algorithm

Dato un modello $\lambda = (A, B, \pi)$ e una sequenza di osservazioni $O = o_1, o_2, \dots, o_T$ si vuole trovare $p(O|\lambda)$. Sebbene possa essere calcolata usando semplici regole probabilistiche avrebbe complessità dell'ordine di $O(N^T)$. Fortunatamente esistono altri metodi di complessità inferiore, uno di questi è il *forward algorithm* che fa uso di una variabili ausiliaria

$$\alpha_t(i) = p(o_1, o_2, \dots, o_t, q_t = i | \lambda) \quad (3.7)$$

dove $\alpha_t(i)$ è detto *variabile forward* e o_1, o_2, \dots, o_t è la sequenza parziale osservata. La relazione ricorsiva è data invece da:

$$\alpha_{t+1}(j) = b_j(o_{t+1}) \sum_{i=1}^N \alpha_t(i) a_{ij}, \quad 1 \leq j \leq N, \quad 1 \leq t \leq T-1 \quad (3.8)$$

con $\alpha_1(j) = \pi_j b_j(o_1)$. Quindi la probabilità richiesta è data da:

$$p(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (3.9)$$

La *variabile backward* può essere calcolata in modo simile.

$$\beta_t(i) = p(o_1, o_2, \dots, o_t | q_t = i, \lambda) \quad (3.10)$$

Avremo quindi due modi di calcolare $p(O|\lambda)$

$$p(O|\lambda) = \sum_{i=1}^N p(O, q_t = i | \lambda) = \sum_{i=1}^N \alpha_t(i) \beta_t(i) \quad (3.11)$$

3.2.2 Viterbi algorithm

Dato un modello $\lambda = (A, B, \pi)$ e una sequenza di osservazioni $O = o_1, o_2, \dots, o_T$ si vuole trovare la sequenza di stati più probabile. L'algoritmo di Viterbi trova la sequenza completa di stati con la massima probabilità. Anche in questo caso è necessario definire una variabile ausiliaria che indichi la massima probabilità di una sequenza di osservazioni parziale, supponendo che i sia lo stato corrente si avrà:

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} p(q_1, q_2, \dots, q_{t-1}, q_t = i, o_1, o_2, \dots, o_{t-1} | \lambda) \quad (3.12)$$

ne consegue che

$$\delta_{t+1}(j) = b_j(o_{t+1}) \left[\max_{1 \leq i \leq N} \delta_t(i) a_{ij} \right], \quad 1 \leq j \leq N, \quad 1 \leq t \leq T-1 \quad (3.13)$$

con $\delta_1(j) = \pi_j b_j(o_1)$, $1 \leq j \leq N$. Questo algoritmo può essere interpretato come una ricerca in un grafo i cui nodi sono formati da stati dell'HMM in ogni istante di tempo t con $1 \leq t \leq T$.

3.2.3 Baum-Welch algorithm

L'algoritmo di Baum-Welch si usa per trovare i parametri sconosciuti di un HMM. Fa uso del ben noto *EM algorithm* per stimare i parametri con la massima probabilità a partire da una sequenza di osservazioni. Fa uso delle procedure di forward e backward già accennate in 3.2.1 ed aggiorna i valori secondo il seguente criterio, facendo uso di due nuove variabili ausiliarie:

$$\xi_t(i, j) = p(q_t = i, q_{t+1} = j | O, \lambda) \quad (3.14)$$

e la probabilità a posteriori

$$\gamma_t(i) = p(q_t = i | O, \lambda) \quad (3.15)$$

Quindi si ha che la relazione tra $\gamma_t(i)$ e $\xi_t(i, j)$ è data da

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j), \text{ con } 1 \leq i \leq N, 1 \leq t \leq M \quad (3.16)$$

Poiché è possibile esprimere sia $\gamma_t(i)$ che $\xi_t(i, j)$ in termini di $a_{i,j}$ e $b_j(k)$ le formule di re-stima saranno le seguenti

$$\overline{a_{i,j}} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \quad 1 \leq i, j \leq N \quad (3.17)$$

$$\overline{b_j(k)} = \frac{\sum_{t=1}^T \xi_t(j, k)}{\sum_{t=1}^T \gamma_t(j)}, \quad 1 \leq j \leq N, 1 \leq k \leq M \quad (3.18)$$

Capitolo 4

Metodo e Remaining Useful Life

In questa tesi abbiamo voluto indagare un metodo unificato che sia diagnostico e prognostico per determinare lo stato di salute di un componente fisico e la sua RUL. Nello specifico questo metodo si basa su controlli non distruttivi ovvero utilizza i dati forniti dai sensori che monitorano le condizioni del componente in esame. I segnali acquisiti vengono prima elaborati per estrarne le caratteristiche e poi utilizzati per apprendere il modello.

4.1 Metodo basato su WPD e HMM

Come appena descritto, i segnali acquisiti vengono dapprima elaborati per estrarne le caratteristiche, questo primo passo verrà eseguito grazie alle WPD e le caratteristiche descritte quindi sotto forma di coefficienti WPD, che verranno poi utilizzati per apprendere il modello di comportamento della degradazione espresso come un HMM. Si noti che nei normali modelli HMM la permanenza in uno stato è una funzione di decadenza geometrica ma che esistono varianti di questo modello che non sono così restrittive.

Il metodo qui proposto si suddivide in due fasi principali, come mostrato

in Fig 4.1: una fase di apprendimento (*learning phase*) off-line, e una fase operativa (*prognostic phase*) on-line.

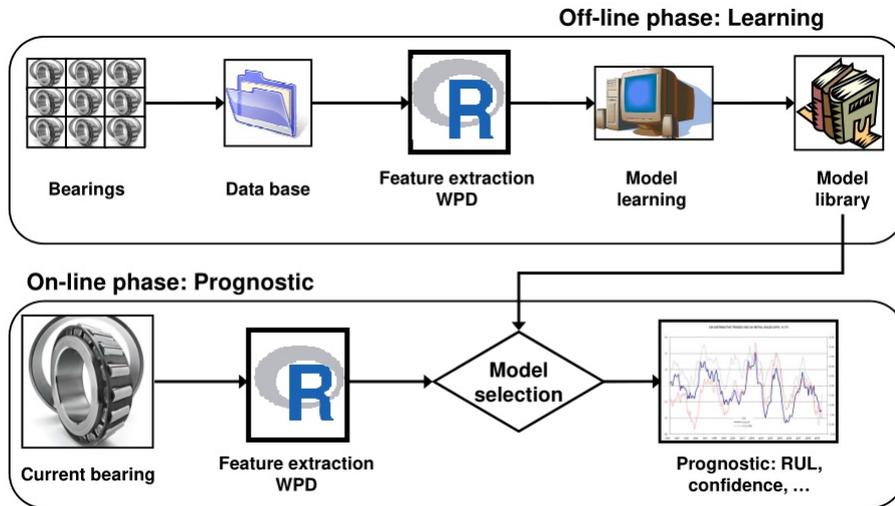


Figura 4.1: Modello generale dei processi della fase di *learnig* e di *prognostic*.

Nella prima fase i dati grezzi rilevati dai sensori vengono elaborati per estrarre l'energia di ciascun nodo rispetto all'ultimo livello di decomposizione, stimato osservando i risultati di decomposizione ai vari livelli utilizzando la tecnica WPD. Queste caratteristiche sono poi utilizzati per apprendere i diversi modelli comportamentali (in forma di HMM), corrispondenti a diversi stati iniziali e condizioni di funzionamento del componente in esame. Infatti ogni serie di dati grezzi corrisponde ad una condizione di un determinato componente e viene trasformata in una matrice funzione F in base a WPD. Nella matrice F , ciascun vettore colonna (caratteristiche di C al tempo t) corrisponde ad un'istantanea sul segnale grezzo, e ciascuna cella f_{ct} rappresenta il nodo c dell'ultimo livello WPD al tempo t .

$$\text{Segnale grezzo} \xrightarrow{\text{WPD}} F = (f_{1t} f_{2t} \dots f_{ct}) \quad (4.1)$$

4. METODO E REMAINING USEFUL LIFE

$$\text{con } 1 \leq t \leq T \text{ e } 1 \leq c \leq C$$

Le energie nodali (caratteristiche) vengono poi utilizzate per stimare i parametri (π, A, B) , ed i parametri temporali (tempo di permanenza in ogni stato) dell'HMM.

Il vantaggio di utilizzare diverse caratteristiche anziché una sola è che una singola caratteristica non può acquisire tutte le informazioni relative al comportamento del componente. I parametri π, A e B di ogni HMM sono appresi utilizzando il noto algoritmo Baum-Welch, mentre i parametri temporali sono stimati utilizzando l'algoritmo di Viterbi. Inoltre quest'ultimo ci permette di ottenere la sequenza degli stati e di sfruttarla per calcolare il tempo di permanenza del componente in ogni stato del corrispondente HMM (Fig. 4.2). Pertanto, assumendo che la permanenza in ogni stato segua una legge gaussiana, è possibile stimare la durata media (4.2), e la deviazione standard corrispondente (4.3) calcolando la durata e il numero di visite in ciascuno stato. Inoltre, l'algoritmo di Viterbi ci permette di identificare lo stato finale, che rappresenta la situazione guasto del componente fisico.

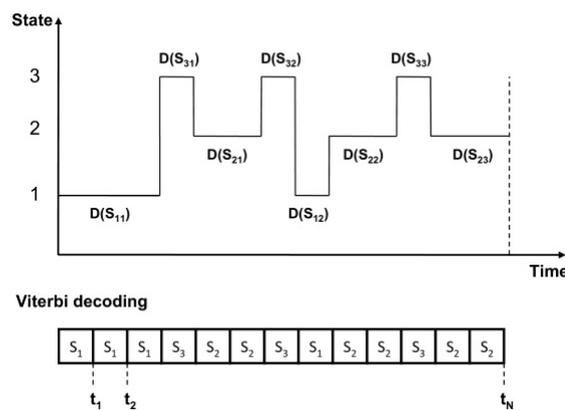


Figura 4.2: Esempio di codifica degli stati di Viterbi.

$$\mu(D(S_i)) = \frac{1}{\Omega} \sum_{w=1}^{\Omega} D(S_{iw}) \quad (4.2)$$

$$\sigma(D(S_i)) = \sqrt{\frac{1}{\Omega} \sum_{w=1}^{\Omega} [D(S_{iw}) - \mu(D(S_i))]^2} \quad (4.3)$$

In (4.2) e (4.3), $D(\cdot)$ sta per la durata della visita, i è l'indice dello stato, w è l'indice di visita, e Ω corrisponde al totale delle visite. Una rappresentazione compatta di ciascun HMM appreso ed utilizzato per eseguire la diagnostica e la prognostica è dato dall'espressione

$$\lambda = (\pi, A, B, \mu(D(S_i)), \sigma(D(S_i)), S_{final}) \quad (4.4)$$

dove S_{final} è lo stato finale (corrispondente alla fine del monitoraggio considerato), $\mu(D(S_i))$ è la durata media per lo stato i , e $\sigma(D(S_i))$ è la deviazione standard della durata per lo stato i .

La seconda fase, che viene eseguita on-line, consiste nello sfruttare i modelli appresi per rilevare la condizione corrente del componente (utilizzando l'algoritmo di Viterbi), e per calcolare la corrispondente RUL. I dati trattati e le energie nodali estratte (utilizzando il pacchetto *wavelets* di R) sono quindi forniti in continuo ai modelli imparati per selezionare quello che meglio rappresenta i dati osservati e quindi le condizioni del componente corrispondente.

Il processo di selezione del modello si basa sul calcolo della probabilità $P(O|\lambda)$ del modello e sulle osservazioni. Infine, conoscendo la condizione attuale e utilizzando le durate in ciascuno stato apprese nella fase di off-line, si possono stimare la RUL del componente e il suo valore di confidenza associato. La stima del RUL viene fatta secondo i seguenti passaggi:

4. METODO E REMAINING USEFUL LIFE

- Il primo passo consiste nel rilevare l'appropriato HMM che meglio si adatta e rappresenta la sequenza osservata di energie nodali. Le caratteristiche sono continuamente ricalcolate sulla base dei modelli appresi e viene calcolato il rischio relativo alla selezione del modello appropriato (Fig. 4.3). Il modello selezionato viene poi utilizzato per calcolare la RUL.

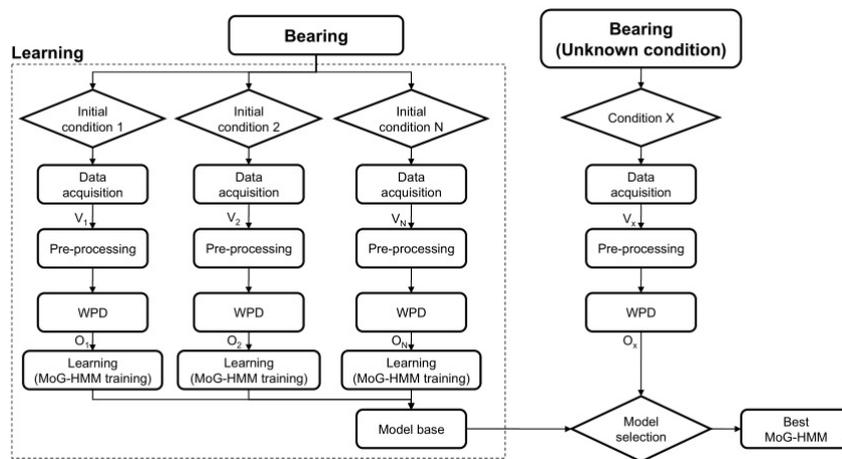


Figura 4.3: Selezione del modello competitivo.

- Il secondo passo di questa procedura riguarda l'individuazione dello stato corrente del componente. L'algoritmo di Viterbi è quindi applicato al modello selezionato per trovare prima la sequenza di stati che corrisponde alla sequenza osservata di energie nodali e poi per identificare lo stato corrente del componente, scegliendo lo stato più persistente nelle ultime osservazioni (S_1, S_2, \dots, S_t) .
- Il terzo passo consiste nell'utilizzare l'attuale stato identificato, lo stato finale (ovvero lo stato che rappresenta il fallimento) e la matrice di transizione di probabilità A dell'HMM selezionato per trovare il *percorso*

critico che va dallo stato corrente allo stato finale. L'idea è di identificare tutte le probabilità non-nulle nella matrice di transizione come potenziali transizioni e quindi di scegliere il percorso minimo tra tutti quelli possibili (Fig. 4.4) con una sola visita per ogni stato. Allo stesso modo, è possibile trovare il percorso più lungo prendendo in considerazione un numero massimo di stati nel percorso con una sola visita per ogni stato. Il percorso più breve è assimilato al percorso pessimistico (rapida evoluzione al fallimento), mentre il percorso più lungo è considerato lo scenario ottimistico.

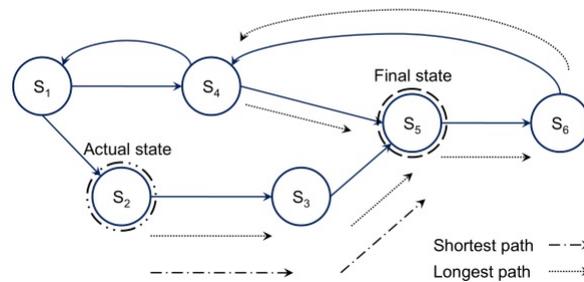


Figura 4.4: Stima del percorso.

- Infine nella quarta fase, i percorsi precedentemente identificati sono utilizzati per stimare la RUL. Essa si ottiene utilizzando i parametri temporali di permanenza in ogni condizione. Inoltre, il valore di confidenza sul RUL viene calcolato in base ai valori di deviazione standard delle durate di permanenza.

$$RUL_{upper} = \sum_{i=currentstate}^N [\mu(D(S_i)) + n \cdot \sigma(D(S_i))], \quad (4.5)$$

$$RUL_{mean} = \sum_{i=currentstate}^N \mu(D(S_i)), \quad (4.6)$$

4. METODO E REMAINING USEFUL LIFE

$$RUL_{lower} = \sum_{i=currentstate}^N [\mu(D(S_i)) - n \cdot \sigma(D(S_i))], \quad (4.7)$$

$\forall i \in stato\ nel\ percorso$, con $n = coefficiente\ di\ confidenza$.

Capitolo 5

WPD e Fault Detection

La WPT si può applicare anche nel problema di *fault detection* preso in considerazione in questa tesi, per elaborare i dati. Ricordiamo che lo scopo è determinare il tempo di vita rimanente a un cuscinetto rotante, sulla base del monitoraggio degli impulsi di vibrazione del segnale. Ampiezza e periodo di questi impulsi sono determinati principalmente dalla velocità di rotazione, dalla localizzazione del guasto e dalle dimensioni del cuscinetto. Ogni guasto avrà quindi una sua frequenza specifica. È noto dalle evidenze sperimentali che il periodo di maggior affaticamento del cuscinetto rotante si ha in corrispondenza dell'accumulo del danno materiale infatti il periodo di propagazione della rottura e del suo sviluppo è molto breve. Questo significa che se usassimo un tradizionale metodo CBM a soglie, il tempo di risposta a disposizione dei manutentori sarebbe veramente breve. Il nostro obiettivo è quello di individuare il processo che ci permette di identificare al meglio il guasto e di farlo nel modo più generale possibile, così da poter applicare le nostre considerazioni al maggior numero di casi.

5.1 Analisi dei dati

In questa sezione vedremo i passi preliminari compiuti per analizzare i dati grezzi a disposizione.

5.1.1 Comprensione di dati

I dati analizzati in questa tesi sono stati generati dal NSF/UCR *Center for Intelligent Maintenance Systems* (IMS - www.imscenter.net). Le misurazioni si riferiscono a quattro cuscinetti installati su un asse in cui la velocità di rotazione è mantenuta costante a 2000 giri al minuti da un motore a corrente alternata accoppiato all'asse mediante cinghie.



Figura 5.1: Foto di un cuscinetto ZA-2115.

I cuscinetti a doppia riga Rexnord ZA-2115 sono installati sull'asse come mostrato in figura 5.2. Tali cuscinetti hanno 16 sferette per ogni riga, il diametro primitivo è di circa 71,5 mm mentre il diametro delle sferette è di circa 8,4 mm (vedi Figura 5.1).

Gli accelerometri al quarzo ad alta sensibilità (ICP) sono stati installati nella sede dei cuscinetti. Abbiamo a disposizione tre data set, nel *data set 1* sono installati due accelerometri per ogni cuscinetto che permettono di misurare le accelerazioni sull'asse x e sull'asse y mentre i *data set 2* e *3* riportano i dati relativi ad un solo accelerometro per cuscinetto. In tutti e tre i casi le

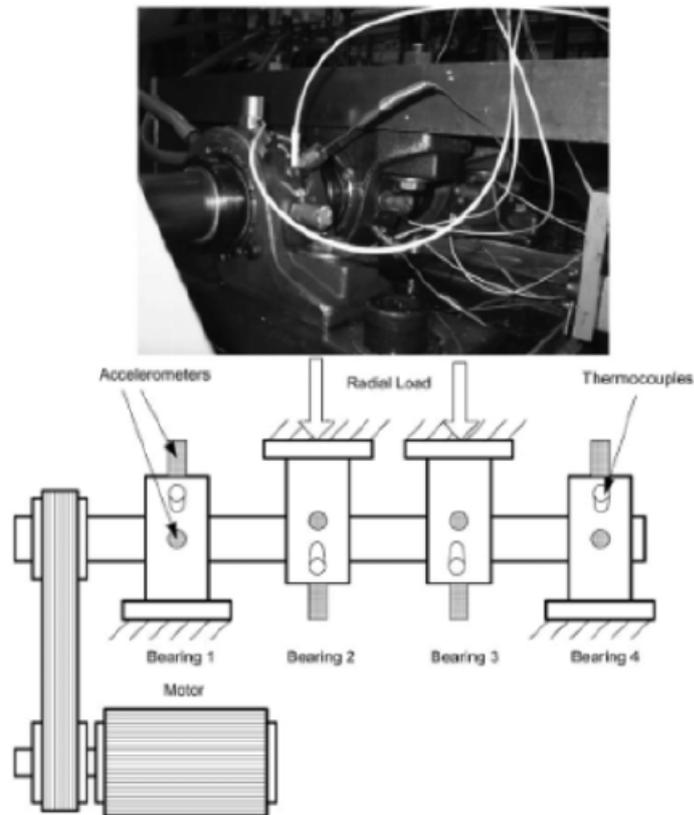


Figura 5.2: Illustrazione della posizione di cuscinetti e sensori.

misurazioni si interrompono quando almeno uno dei cuscinetti incorre nel fallimento, ovvero quando sarebbe necessario sostituirlo per poter proseguire il lavoro. Tutti i fallimenti si sono verificati dopo aver superato il tempo di vita dichiarato per il cuscinetto che è maggiore di 100 milioni di giri.

Data set 2

Nello specifico, in questa tesi ci siamo occupati di analizzare il *data set 2* in quanto presenta dati con caratteristiche più evidenti.

Il *data set 2* consiste di 984 file ottenuti ad intervalli regolari di 10 minuti l'uno dall'altro per un periodo di circa sette giorni. All'interno di ciascun file

si trovano 20048 righe rappresentanti un'istantanea della durata di un secondo delle vibrazioni del cuscinetto e quattro colonne, una per ogni cuscinetto del sistema di figura 5.2. Le misurazioni terminano per il fallimento avvenuta nel cuscinetto 1 (prima colonna) che è di tipo *ouerer race*. Le vibrazioni sono interpretabili come scostamento dall'asse di riferimento perciò assumono valori negativi e positivi.

5.1.2 Selezione dei dati

Si è scelto di utilizzare come tool di analisi R ed in particolare il pacchetto *wavethresh* che permette di calcolare agevolmente la WPD. La funzione richiede in ingresso una struttura di dati che abbia lunghezza pari a una potenza di 2. Poiché $2^{14} = 16384$ è la più grande potenza di 2 contenuta in 20048 si è scelto di dare in input le prime 16384 righe di ogni file. Questo non implica perdita di generalità ma solo che ciascuna osservazione sia più breve di un secondo.

5.1.3 Preparazione ambiente di calcolo

Essendo la lunghezza dei dati pari a 2^{14} la WPD avrà 14 livelli, nella notazione del programma il 14° livello è quello che non è ancora stato trasformato ovvero la *mother wavelet* ed ha un solo *index* indicato con il valore 0. Poiché possiamo considerare almeno i livelli 14 e 13 di scarso interesse a causa della scarsa decomposizione via wavelet, la nostra analisi prenderà in considerazione i livelli a partire dal 12, a scendere.

Per poter visualizzare graficamente dove siano i picchi che distinguono le tre fasi: sano, danno lieve e fallito; è fisicamente sensato calcolare dai dati l'energia 5.1 o la media dei valori assoluti 5.2.

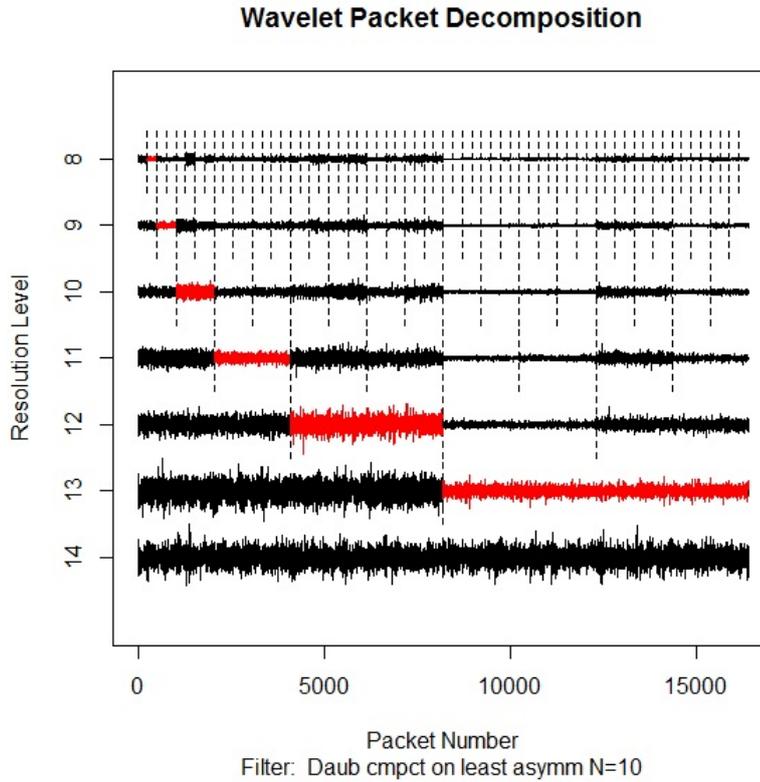


Figura 5.3: WPD relativa al test2 con livelli a partire da 8

$$E = \left(\frac{\sum_{k=1}^M x_k^2}{M} \right)^{\frac{1}{2}} \quad (5.1)$$

$$M = \frac{\sum_{k=1}^M |x_k|}{M} \quad (5.2)$$

dove x_k è un valore dei dati di oscillazione e M è il numero dei campioni nel nodo. Vanno quindi inizializzate, per ogni livello che si vuole studiare, due matrici: una per l'energia e una per le medie dei valori assoluti. Tali matrici avranno tante righe quanti sono gli *index* del livello selezionato e tante colonne quanti sono i file nel test.

5.1.4 Osservazioni preliminari

Osservando la Figura 5.4 possiamo notare che i grafici ottenuti con i valori dell'energia di decomposizione e quelli ottenuti con la media dei valori assoluti, portano a conclusioni identiche, quindi in seguito ci si potrà concentrare esclusivamente sull'energia essendo un indice più comune in letteratura per descrivere i livelli delle wavelets.

Un'altra considerazione fatta sul primo test e valida per ridurre i confronti successivi è la seguente: riassumendo tutti i dati ottenuti da un test in una tabella come in Figura 5.5 si nota che il livello 11 è l'ultimo livello in cui i punti non si raddoppiano, è quindi quello di maggior interesse.

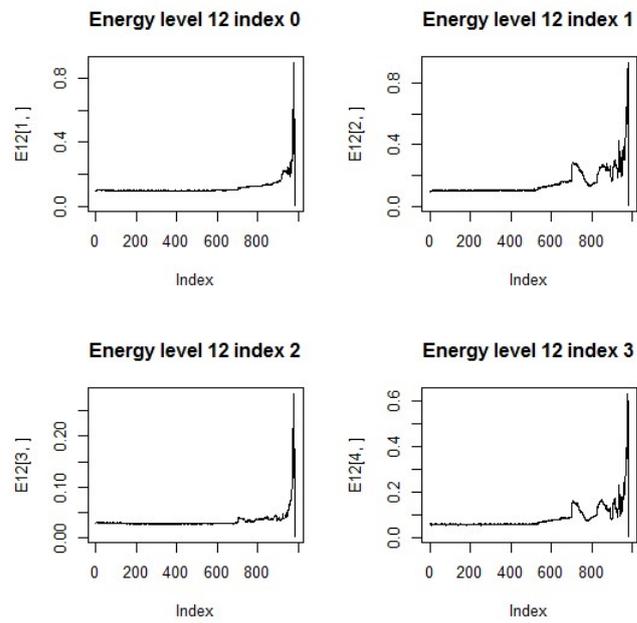
5.1.5 Interpretazione grafica

Nella Figura 5.6 sono riportati i grafici relativi all'energia al livello 11 per index 2 e 6, ovvero quelli ritenuti più significativi nel senso che è maggiormente evidente la suddivisione in stati.

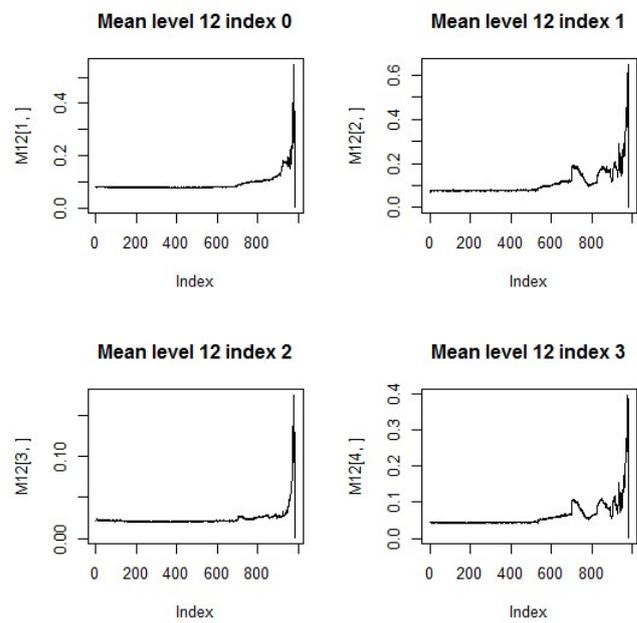
Da essi possiamo dedurre, relativamente al test2, che sia opportuno dividere gli intervalli nel seguente modo:

- 0 – 500 sano
- 500 – 800 danno lieve
- 800 – 984 fallito

5. WPD E FAULT DETECTION



(a) Energy



(b) Mean

Figura 5.4: Esempio di energia e media dei valori assoluti.

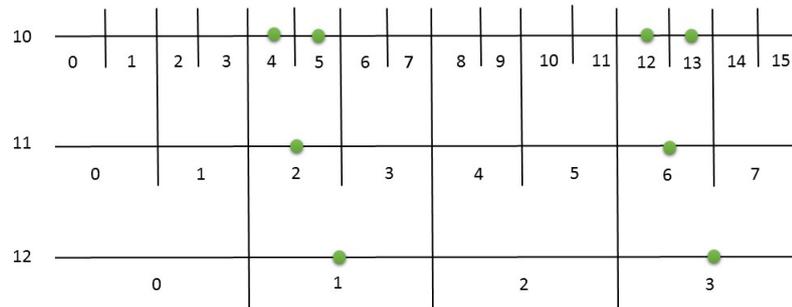
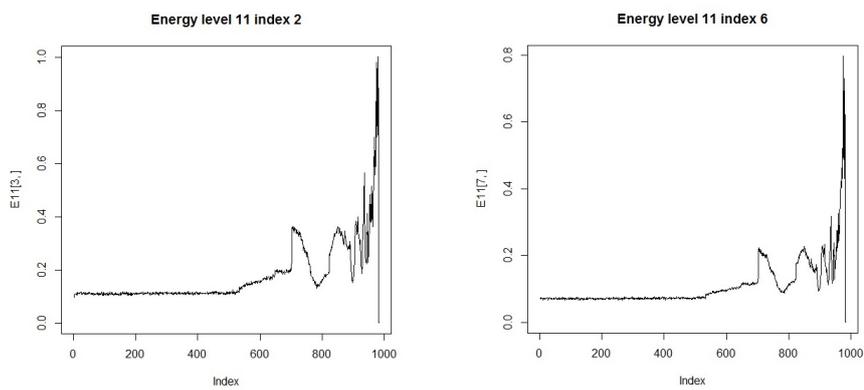


Figura 5.5: Tabella riassuntiva, in verde gli indici più interessanti.



(a) level 11 index 2

(b) level 11 index 6

Figura 5.6: Grafici dell'energia per il livello 11 del test2.

Capitolo 6

Applicazione di HMM

Una volta elaborati i dati con la trasformata wavelets e scelta la vibrazione caratteristica con un opportuno livello e index (nel nostro caso livello 11, index 2) abbiamo a disposizione un campione rappresentativo di tutto l'intervallo temporale ma che porta meno informazione, infatti è stato filtrato ed è distribuito come in Figura 6.1.

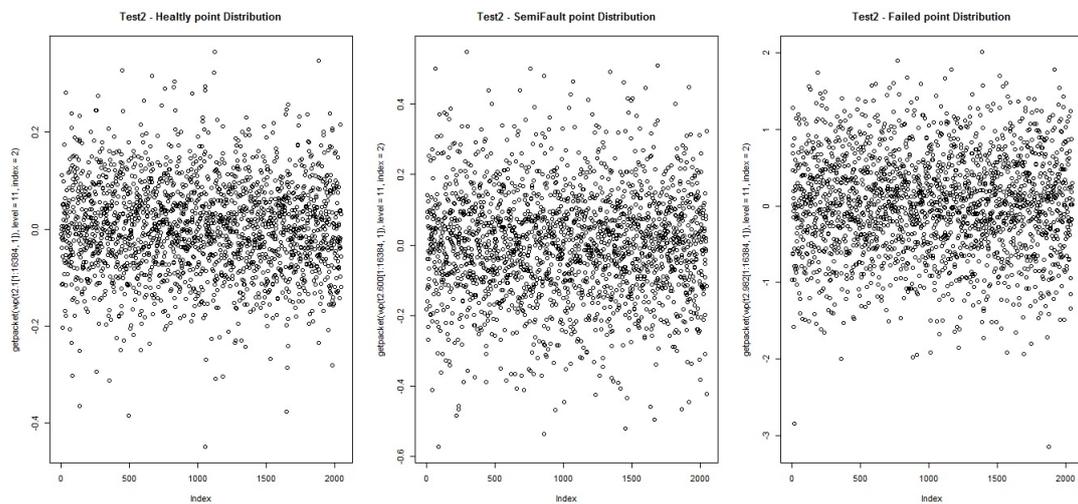


Figura 6.1: Esempio di distribuzioni dei punti di 3 secondi consecutivi.

Per prima cosa, è necessario costruire un HMM, in una prima analisi lo co-

struiremo con tre stati: Healthy (H), Semi-Failed(SF) e Failed (F) (vedi Figura 6.2). Successivamente vedremo come varia il modello se gli stati considerati sono quattro. I passaggi concettuali sono gli stessi quindi nel seguito faremo riferimento esclusivamente ai tre stati per descrivere le procedure seguite senza perdere di generalità.



Figura 6.2: Stati della Markov chain di partenza.

Il risultato della wavelets transform è un insieme di osservazioni espresse come vibrazioni. In questa forma quindi non sono evidenti gli stati in cui si trova il sistema. Per farli emergere è possibile costruire un *Hidden Markov Model* come in Figura 3. Ovvero gli stati *hidden* sono "Healthy", "Semi-Failed" e "Failed" mentre le osservazioni sono rappresentate dai pallini sottostanti la catena di Markov e sono noti.

Nello specifico, il sistema qui considerato è discreto nel tempo ma continuo nello spazio. Per trattare questo tipo di problemi, ovvero con dati continui, abbiamo due possibili approcci:

- discretizzare lo spazio associando ad ogni valore uno stato fittizio
- trovare e sfruttare la distribuzione relativa ad ogni stato (schematizzato in Figura 6.4)

In questo studio si è scelta la seconda strategia perché oltre ad offrire risultati migliori [5] permette anche una maggiore generalizzazione.

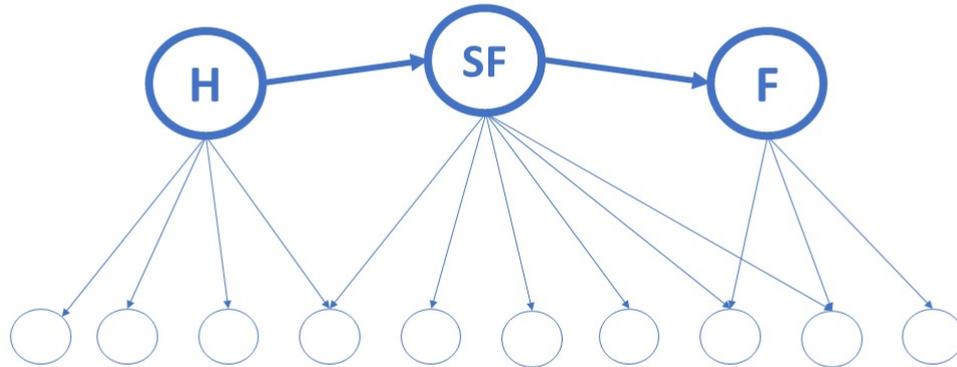


Figura 6.3: Esempio di Hidden Markov Model a 3 stati.

Ovviamente prima di poter applicare questa strategia è bene verificare che i dati di partenza abbiano effettivamente una forma normale. Oltre ai ben noti test statistici di normalità è possibile verificare la normalità dei nostri dati anche attraverso l'osservazione degli istogrammi relativi ai tre stati (esempio in figura 6.5). Risulta quindi che la distribuzione degli stati è di tipo gaussiano e che i parametri caratteristici della curva sono diversi e quindi caratterizzanti per ciascuno stato. Fatta questa verifica si può procedere con l'algoritmo che ci permette di creare un modello per il nostro problema e sfruttarlo per individuare la sequenza ottimale.

Dalle sperimentazioni è emerso che è più significativo e anche più agevole considerare l'energia relativa ad ogni istante osservato piuttosto che ogni singola vibrazione e quindi le computazioni seguenti si baseranno su tale dato.

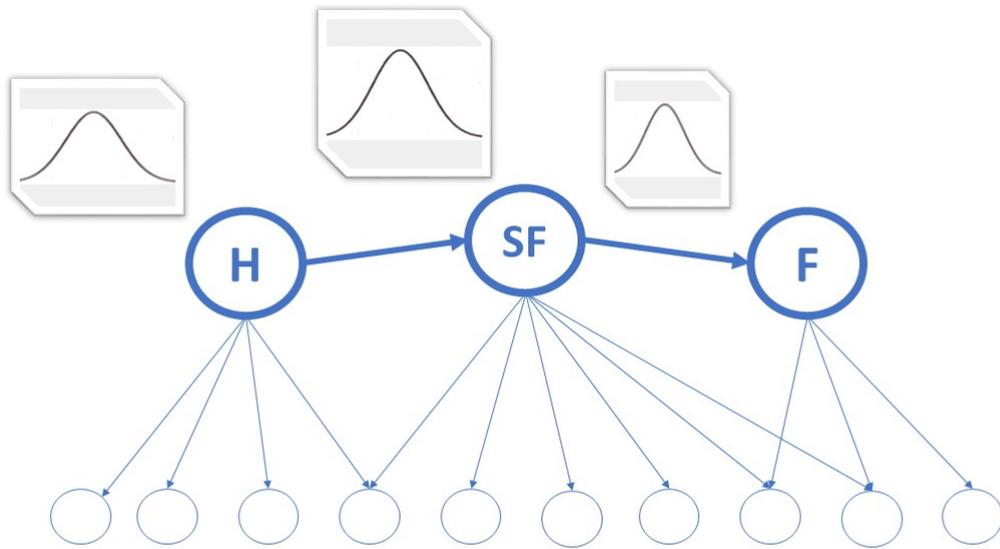


Figura 6.4: Esempio di Hidden Markov Model a 3 stati ognuno con associata una specifica distribuzione di forma gaussiana.

6.1 Ottimizzazione

Per poter ottenere il modello siamo partiti dalla seguente quaterna iniziale (P_i, δ, m, σ) utilizzata come input per l'algoritmo Baum-Welch.

$$P_i = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \end{pmatrix} \quad (6.1)$$

La matrice P_i (6.1) rappresenta la matrice delle transizioni di probabilità tra gli stati

$$\delta = (1, 0, 0) \quad (6.2)$$

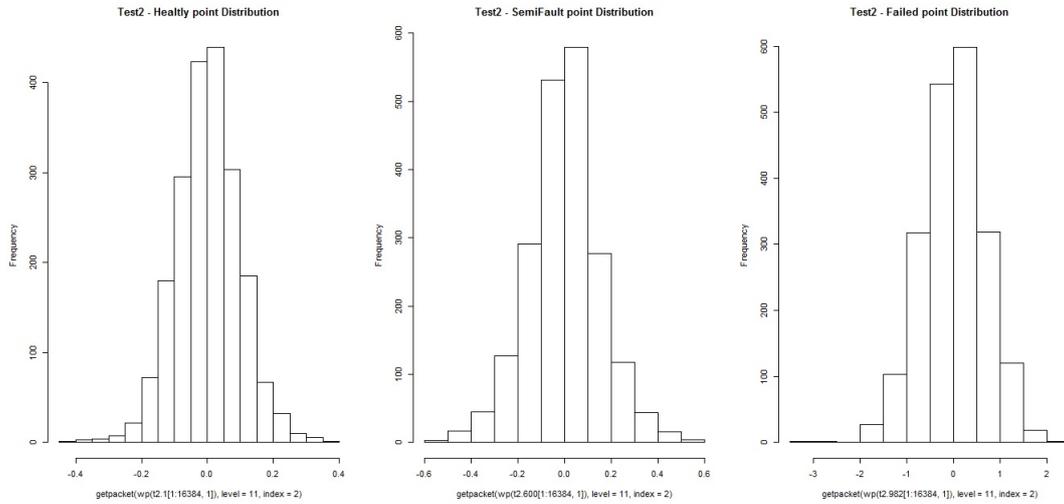


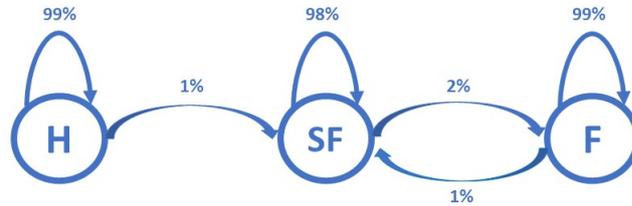
Figura 6.5: Istogrammi delle distribuzioni relativi agli stati H, SF,F per il test3.

Il vettore δ (6.2) rappresenta il fatto che supponiamo di partire sempre da uno stato *Healthy*. m è il vettore contenente le medie caratteristiche dei tre stati mentre σ sono le rispettive varianze delle distribuzioni.

A partire dai dati e da questi parametri iniziali, l'algoritmo Baum-Welch è in grado, in poche iterazioni, di restituire un modello (vedi Figura 6.6) sufficientemente accurato.

6.2 Sequenza ottimale

Il problema di trovare la sequenza ottimale di stati associata alle osservazioni si basa in grande misura sul significato che si sceglie di attribuire a *ottimale*. Uno dei possibili criteri è quello di scegliere lo stato q_t che sia individualmente il più probabile. Con questa scelta però si può incorrere nel problema dell'annullamento infatti se una delle transizioni tra stati ha probabilità zero,



(a) 3 stati



(b) 4 stati

Figura 6.6: Modello risultante dall' algoritmo Baum-Welch per il test 2 considerando (a) 3 stati e (b) 4 stati.

la sequenza ottimale può essere invalidata. Questa soluzione infatti determina lo stato più probabile in ogni istante però non tiene conto della probabilità

dell'occorrenza delle sequenze di stati. Una possibile soluzione potrebbe essere quella di trovare la sequenza di stati che massimizza il numero atteso di coppie di stati (o triple, ecc.). Ad ogni modo la tecnica più comune è quella di massimizzare $P(Q, O|\lambda)$ dove $Q = q_1, q_2, \dots, q_T$ è la sequenza ottimale e $O = O_1, O_2, \dots, O_T$ è la sequenza di osservazioni data. Una delle tecniche formali più diffuse, basata su metodi di programmazione dinamica, è l'*algoritmo di Viterbi* (vedi sezione 3.2.2).

Capitolo 7

Step per il calcolo di RUL

Per poter calcolare la RUL dobbiamo fare ancora alcuni passaggi a partire dal modello. Fino ad ora abbiamo ottenuto l'HMM e abbiamo sfruttato l'algoritmo di Viterbi come verifica visualizzando la sequenza ottimale e confrontandola con quella da cui siamo partiti ottenendo così il numero di errori commessi dal modello. Con l'algoritmo di Viterbi però, possiamo anche ottenere la sequenza ottimale di una qualsiasi sotto sequenza di dati, dal risultato di questo, quindi, possiamo ricavare lo stato in cui si trova il sistema in ciascun istante ovvero fare un monitoraggio *online*. Per farlo estraiamo dal risultato di Viterbi lo stato che compare con maggiore frequenza tra gli ultimi l termini (secondo la notazione di Tobon-Mejia e colleghi[5]).

7.1 Verifica dell'accuratezza

Poiché i dati a nostra disposizione riguardano un solo esperimento è stato necessario suddividerli in training e test set. Per verificare l'accuratezza del nostro modello utilizziamo il test set come descritto qui di seguito. Poiché, per ipotesi, avremo una fase online è sensato pensare di suddividere il test set

in sotto sequenze con un fissato numero di item (misurazioni di 1 secondo) aggiungendo ogni volta un novo dato in coda e togliendone uno in testa. Per chiarificare il concetto supponiamo che n sia il numero totale di item nel test set e sia s il numero di item di cui è composta ogni sotto sequenza di item. Allora suddividiamo il test set come segue:

$$[(x_1, \dots, x_s), (x_2, \dots, x_{s+1}), \dots, (x_{n-s}, x_n)] \quad (7.1)$$

dove x_i rappresenta l' i -esimo item.

Per scegliere al meglio i valori di s ovvero l'ampiezza delle sotto sequenze di item e di l cioè su quanti valori finali basare l'attribuzione di uno stato a un sotto sequenza abbiamo sfruttato la funzione d'errore accennata nell'introduzione di questo capitolo. Tale funzione calcola quante volte il nostro modello è concorde con il risultato atteso. Per descrivere il calcolo dell'errore in modo formale facciamo riferimento alla suddivisione degli item descritta in 5.1.5 che rappresenta il risultato atteso. Gli item del primo intervallo sono arbitrariamente considerati sani (stato = 1), quelli del secondo con danni lievi (stato = 2) mentre quelli del terzo con fallimento del sistema (stato = 3). Siano h , sh e f le rispettive soglie massime dei suddetti intervalli. Sia poi $[x_a, \dots, x_b]$ una sotto sequenza e sia t il numero totale di elementi nel dataset considerato. Allora la funzione d'errore è esprimibile con il seguente pseudocodice:

```

for i=1 to (t-s)
  calcolo l'HMM della i-esima sotto sequenza
  v = il valore (1, 2, 3) dello stato della sotto sequenza
    i-esima calcolato con l'algoritmo di Viterbi
  if (b>0 && b<=h && v!=1) error<-error+1
  if (b>h && b<=sf && v!=2) error<-error+1
  if (b>sf && b<=f && v!=3) error<-error+1

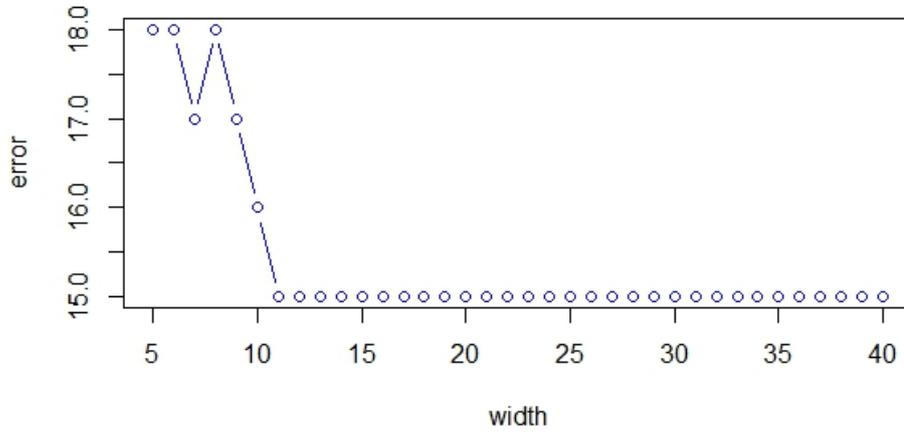
```

7. STEP PER IL CALCOLO DI RUL

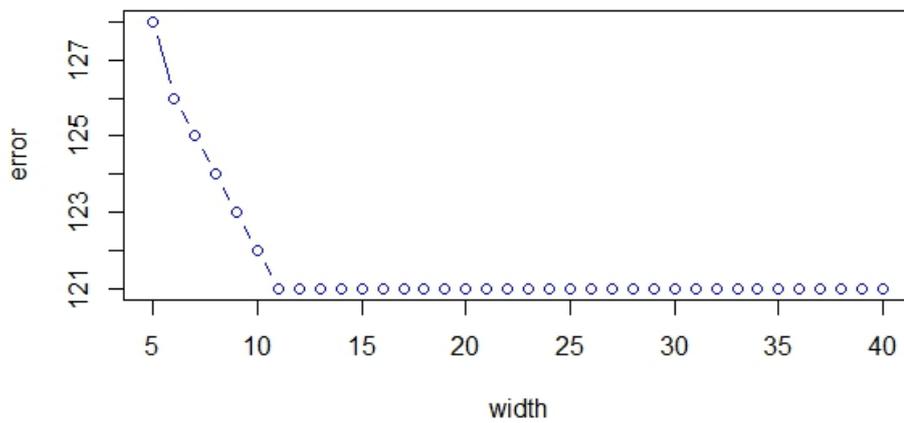
```
a=a+1  
b=b+1  
end
```

Quindi sfruttando tale funzione considerando via via tutti i possibili valori di l per una fissata sotto sequenza di item di lunghezza s si evince che l'errore minimo commesso si ha in presenza dell'uso del solo ultimo elemento della sequenza, ovvero nella stima dello stato corrente si minimizza l'errore se si considera solo l' s -esimo elemento di ogni sotto sequenza. Con un indagine analoga effettuata invece su tutte le possibili lunghezze s di sotto sequenze si può mettere in relazione l'ampiezza dell'intervallo s e l'errore commesso (vedi Figura 7.1).

Osservando le figure si può notare che l'errore ha un andamento non lineare ma con asintoto orizzontale per il valore di errore pari a 11. Per via del comportamento asintotico tutti i valori di s da una certa soglia in poi commettono lo stesso numero di errori e quindi sotto questo aspetto sono equivalentemente validi. Tuttavia le evidenze sperimentali hanno evidenziato che sia preferibile scegliere l'intervallo di ampiezza minore ai fini di minimizzare l'errore relativo medio commesso dal modello. Allo scopo di chiarificare il concetto si può notare, relativamente alla Figura 7.1a che scegliendo l'ampiezza 11, la minore a cui corrisponde errore minimo la percentuale di errore è di circa il 3% mentre scegliendo qualsiasi altro intervallo di maggiore ampiezza la percentuale di errore aumenta perché diminuisce il numero di intervalli in cui è possibile suddividere il test set. Questo banale ragionamento non può essere trasportato così facilmente in un processo online, cioè non è sempre vero che l'ampiezza $s = 11$ commetta il 3% di errore perché non si ha a disposizione il totale delle osservazioni ma rimane comunque vero il fatto che l'errore sarà minore in presenza di sequenze più corte, nell'esempio di sequenze lunghe 11 *item*.



(a) 3 stati



(b) 4 stati

Figura 7.1: Il grafico mostra l'errore in funzione dell'ampiezza s scelta, a seconda che abbia (a) 3 stati oppure (b) 4 stati.

Una volta stabiliti questi parametri è possibile iniziare lo step del calcolo della RUL vero e proprio.

7.2 Calcolo della RUL

Per poter stimare la RUL dobbiamo analizzare accuratamente come sono stati raccolti i dati, in particolar modo se le misurazioni sono state regolari e con quale regolarità. Nel caso del test2 qui analizzato i file sono presi in modo regolare con cadenza di 10 minuti per un totale di circa 6 giorni. Il primo passo da compiere per poter stimare la RUL è quello di ottenere la durata in ciascuno stato. Poiché il modello ottenuto ha anche probabilità di tornare indietro (vedi Figura 6.6) dobbiamo considerare la media e la varianza della permanenza in ciascuno stato.

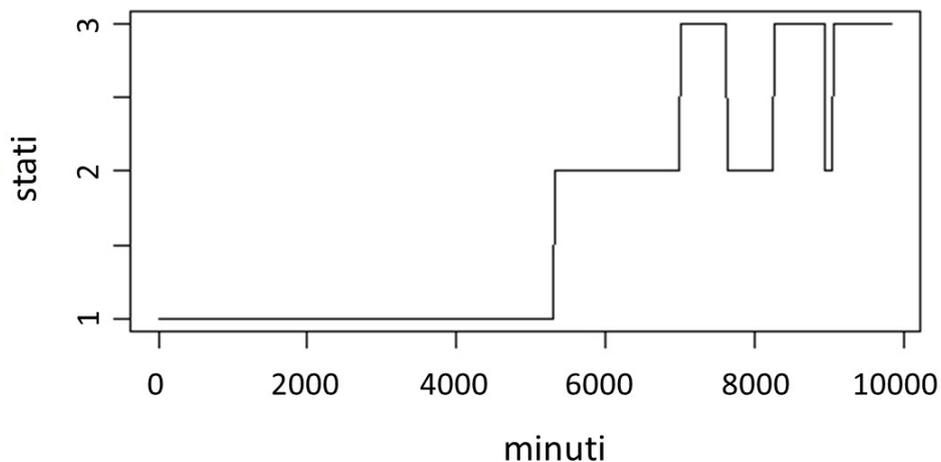


Figura 7.2: Codifica degli stati risultanti da Viterbi per il test2.

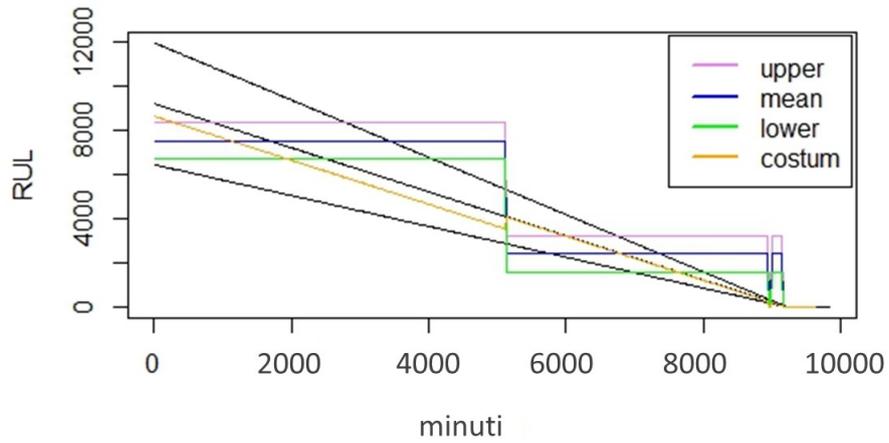
A partire dalla media ottenuta dal grafico si calcola la RUL relativa a ciascuno stato moltiplicando la permanenza nello stato per la durata effettiva di ciascun salto temporale tra un item e l'altro.

Nella fase online, a partire dal risultato dell'algoritmo di Viterbi, dove con risultato intendiamo lo stato in cui si trova il sistema data una certa sotto sequenza, possiamo stimare la RUL con i dati di media e varianza secondo diversi tipi di stima:

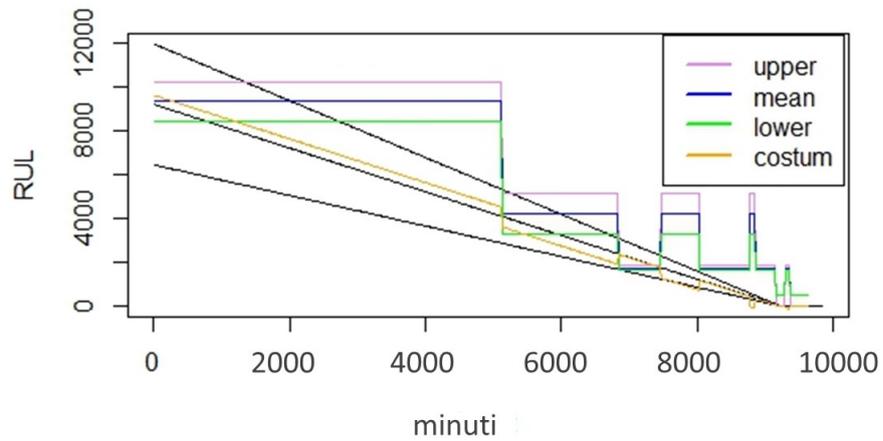
- stima ottimistica;
- stima in media;
- stima pessimistica;
- stima custom.

Le prime tre stime, ottimistica, media e pessimistica, sono state già descritte e si riferiscono dell'ordine alla stima fatta considerando la durata media più la relativa varianza, la durata media e la durata media meno la relativa varianza. In questo studio si è scelto di moltiplicare la media per il numero di volte in cui compare lo stato ottenendo così una stima più verosimile all'effettiva durata in uno stato. Questa scelta è parzialmente in disaccordo con il modello che mostra la possibilità di tronare a uno stato precedente ma è in accordo con la realtà fisica di un sistema meccanico dove, quindi, dallo stato di guasto/fallimento non si può tornare autonomamente a uno stato precedente.

Per quanto riguarda la stima che qui chiamiamo *custom* si è tenuta in considerazione la formula, tra le tre precedenti, che meglio rappresenta il problema e l'informazione di quanto tempo sia già durato ciascuno stato sottraendola al tempo atteso in modo da ottenere un andamento più simile al lineare scorrimento del tempo. Ad esempio nel caso del test 2 a tre stati si è considerata la formula della stima ottimistica mentre nel caso a quattro stati quella media. Infatti dalla Figura 7.3 si può osservare che sono proprio le stime che più si avvicinano alla retta reale della RUL.



(a) 3 stati



(b) 4 stati

Figura 7.3: Grafico dell'andamento delle stime della RUL con tolleranza 30%

7.3 Stima della bontà

Per poter mettere a confronto i vari modelli che possono essere creati su un set di dati si è deciso di far riferimento all'errore assoluto e all'errore relativo.

Per rendere agevole e più significativo il confronto si fa riferimento all'errore relativo medio commesso dalla stima custom. Un altro indice tenuto in considerazione è il coefficiente di log likelihood che rappresenta la probabilità (vedi Appendice A).

Capitolo 8

Conclusioni

Questo esperimento si è concluso con il confronto tra i vari modelli che possono essere generati, con la procedura descritta, a partire dai dati.

8.1 Confronto tra i modelli

Come si può notare dai grafici in Appendice A sono stati considerati per il confronto solo i livelli più significativi: 12, 11, 10. Il livello 12 presenta un livello di decomposizione ancora troppo basso per poter costruire modelli affidabili mentre oltre al livello 10 la decomposizione risulta essere eccessiva. Proprio il livello 10 presenta risultati molto interessanti ma non generalizzabili perché andrebbe scelto l'index valutando caso per caso. Quindi dal confronto si evince che il miglior livello di decomposizione, in senso generale, sia il livello 11.

Per quanto riguarda la scelta del numero di stati su cui costruire il modello si ha, dalle evidenze sperimentali, che il modello a 4 stati è nettamente migliore a livello 10 ma leggermente peggiore a livello 11. La scelta più opportuna in senso generale resta però quella di costruire il modello con il maggior nu-

mero di stati perché a fronte di un errore di poco maggiore nel nostro test si ha una migliore rappresentazione del problema reale.

8.2 Accurateza

Nella procedura descritta, l'errore relativo medio minimo commesso si aggira attorno al 13%. In un'applicazione reale si manifesta sicuramente come un miglioramento rispetto ai precedenti sistemi di CBM. Inoltre un modello più accurato e addestrato su una maggiore quantità di dati potrebbe portare a risultati molto migliori. Da non sottovalutare anche l'uso di tecniche per la costruzione dei modelli accennate nel Capitolo 9.

Capitolo 9

Altri approcci al problema della manutenzione predittiva

Esistono in letteratura molti altri approcci a questo problema, alcuni sono solo varianti altri sono vere e proprie visioni alternative. I risultati di tali approcci non sono paragonati a quello qui sviluppato, in alcuni casi per mancanza di dati coerenti in altri perché proprio i dati di partenza riguardavano diversi dataset. Ad ogni modo si tratta di spunti interessanti per successive trattazioni che possono dare un'idea di insieme sul problema trattato.

9.1 Varianti

In questa sezione vedremo le proposte in letteratura per trattare il problema di questa tesi con diversa analisi dei dati o diversa costruzione del modello basata comunque sul concetto delle catene di Markov.

9.1.1 Diversa decomposizione dei dati

Nel paper [22] si propone di estrarre le caratteristiche effettive delle vibrazioni utilizzando una Dynamic Principal Component Analysis (DPCA). La DPCA è una tecnica per la semplificazione dei dati utilizzata in genere nell'ambito della statistica multivariata lo scopo primario di questa tecnica è la riduzione di un numero più o meno elevato di variabili (rappresentanti altrettante caratteristiche del fenomeno analizzato) in alcune variabili latenti (feature reduction).

9.1.2 Diversa costruzione del modello

Nei paper [5] [31] [30] al posto del normale HMM si fa uso del mixture of gaussian HMM (MoG-HMM) ciò permette di gestire la complessità della densità di probabilità di emissione generate da un insieme di funzioni continue estratte dai segnali di monitoraggio e di rappresentare meglio l'andamento nel tempo. Di solito, le HMM considerano le osservazioni come simboli discreti e utilizzano densità di probabilità discrete per modellare la transizione e le probabilità di osservazione. Il problema di questo approccio è che, nel monitoraggio delle condizioni, le osservazioni sono tipicamente segnali continui. I MoG-HMM consentono invece di modellare il degrado del componente fisico utilizzando osservazioni continue fornite dal monitoraggio sensori. Inoltre permettono di stimare le durate di permanenza in ciascun stato portando alla previsione del valore RUL [33].

9.2 Alternative

In questa sezione vengono brevemente descritti alcuni degli approcci proposti in letteratura per la stima della RUL che considerano il problema con un

approccio diverso da quello seguito in questa tesi.

9.2.1 PAC e LNPP

Nel suo lavoro Jianbo Yu [23] propone l'utilizzo di un algoritmo per l'estrazione delle caratteristiche basato sulla preservazione di proiezioni locali e non locali (LNPP) che differisce dalla Principal Component Analysis (PCA) il cui scopo è scoprire la struttura globale nello spazio Euclideo e regolare gli algoritmi di apprendimento in modo che preservino la struttura locale dei dati. LNPP è in grado di scoprire sia le strutture locali che quelle non locali. Questo di trovare le informazioni significative a bassa dimensionalità nascoste dalle caratteristiche ad alta dimensionalità e quindi serve come preprocesso per la classificazione dei difetti e di conseguenza per la costruzione del modello.

9.2.2 Cross-Entropy e Cross-Validation

Sergey Porotsky, Zigmund Bluvband ALD Group [25] basano il loro studio sull'osservazione del fatto che i dati relativi al guasto di un sistema sono *non-trendability* ovvero non è possibile stabilire dei parametri che ne definiscano significativamente il comportamento evitando così l'uso di modelli di previsione RUL e di apprendimento supervisionato. Creano quindi nuovi modelli ed inoltre utilizzano il metodo Cross-Entropy, ossia la misura di divergenza tra due modelli, per il controllo dell'ottimizzazione dei parametri basata sulla procedura Cross-Validation. Per costruire il loro modello si basano non sulle vibrazioni direttamente misurate ma su valori accumulati, questo viene giustificato fisicamente perché il degrado accumulato in un certo periodo di tempo è proporzionale, in termini di vibrazioni, all'accelerazione accumulata. Propongono quindi due possibili metodi per gestire le vibrazioni:

- Eseguire l'accumulazione degli integrali dei valori della vibrazione invece dei valori di vibrazioni orizzontali e verticali.
- Utilizzare uno dei metodi di apprendimento più diffusi per la previsione RUL (SVR - Regressione Vettoriale di Supporto, RVM - Macchine Vettoriali di Rilevanza, ecc.) per ottenere dati multi parametrici dopo l'esecuzione di accumuli orizzontali e verticali fatti separatamente.

9.2.3 LAD

Nei paper [26] [28] si propone un'analisi dei dati che non presupponga la costruzione di un modello matematico. Tale classificazione dei fenomeni si basa semplicemente sull'estrazione di caratteristiche e l'individuazione di pattern ed è detta LAD (Logical Analysis of Data)

9.2.4 SVM

David Martínez-Rego, Oscar Fontenla-Romero, Amparo Alonso-Betanzos [27] nel 2011 propongono un approccio basato su SVM (Support Vector Machine) seguendo la metodologia qui sotto riassunta:

- estrarre i dati dai possibili punti in cui si verifica il fallimento;
- ridurre la dimensionalità dei dati grezzi;
- usare la SVM per individuare i cambiamenti;
- valutare secondo tre livelli: simulazione, laboratorio, dati reali.
- comparare i risultati mediante ANN (Artificial Neural Network).

La SVM verrà addestrata a distinguere tra situazioni *normali* e *non normali* ovvero a distinguere i guasti in modo binario o booleano.

Appendice A

Dati riassuntivi dei modelli a confronto

I dati delle tabelle rappresentano i coefficienti di loglikelihood per ogni index di ogni level preso in considerazione relativamente alla trasformata wavelet. La likelihood è definita come la densità di probabilità che un set di misure $(x_1, x_2, x_3, \dots, x_N)$ sia prodotto a partire dal particolare valore a .

$$L(x_1, x_2, \dots, x_N; a) = f(x_1; a)f(x_2; a) \cdots f(x_N; a) = \prod f(x_i; a) \quad (\text{A.1})$$

Esprime il fatto che la probabilità che la i -esima misura sia compresa tra x_i e $x_i + dx_i$ è $f(x_i; a)dx_i$ e le N misure sono indipendenti. Quindi dati $(x_1, x_2, x_3, \dots, x_N)$ si può stimare il parametro a che caratterizza $f(x_i; a)$ massimizzando la likelihood.

I dati relativi alla loglikelihood permettono qui un'analisi più mirata per la selezione del modello e giustificano la scelta iniziale che era ricaduta sul level 11 con index 2 a partire dalla mera osservazione dell'andamento dei dati.

I grafici che seguono invece permettono di individuare quanto espresso dalla tabella A.1, ovvero risultati minori, dove con minori si intendono i più

Tabella A.1: Dati relativi alla loglikelihood per i livelli 10, 11, 12 per ogni index.

MODELLO A 3 STATI

conf12	min index	1	max index	2
index	0	1	2	3
LL	-1972,679	-2766,225	-1417,381	-2461,817

conf11	min index	2	max index	4				
index	0	1	2	3	4	5	6	7
LL	-1902,545	-1790,543	-2160,603	-2110,693	-1025,502	-1233,169	-1918,545	-1373,109

conf10	min index	7	max index	8				
index	0	1	2	3	4	5	6	7
LL	-1429,9386	-1604,2009	-1430,7032	-1515,1679	-1901,2777	-1868,0556	-1430,3375	-1957,0003
index	8	9	10	11	12	13	14	15
LL	-716,7787	-808,9365	-1012,8986	-949,2141	-1684,0053	-1824,726	-954,6722	-1188,3005

MODELLO A 4 STATI

conf12	min index	1	max index	2
index	0	1	2	3
LL	-1956,922	-2284,237	-1394,019	-2054,914

conf11	min index	2	max index	4				
index	0	1	2	3	4	5	6	7
LL	-1890,2272	-1769,0121	-2095,2356	-2089,1754	-995,7233	-1215,0079	-1876,7373	-1361,2908

conf10	min index	1	max index	11				
index	0	1	2	3	4	5	6	7
LL	-1550,521	-2428,647	-1700,453	-1748,337	-1911,63	-1867,989	-1932,503	-1782,336
index	8	9	10	11	12	13	14	15
LL	-1182,544	-1344,626	-1473,866	-1204,886	-1850,626	-1658,851	-1410,817	-1438,632

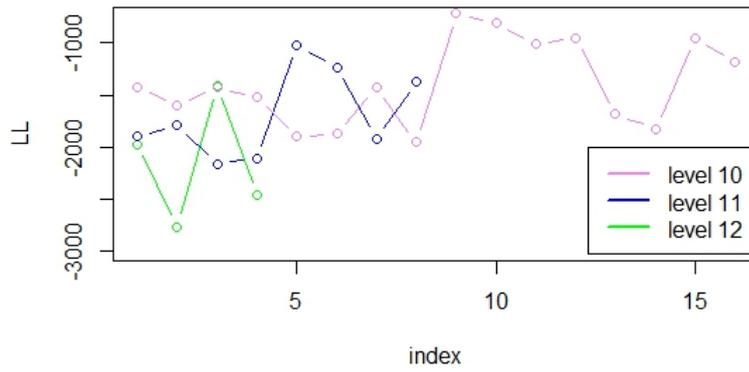
lontani dallo zero, indicano previsioni migliori e viceversa. Alcuni dei punti però non sono significativi per via delle difficoltà ad individuare un adeguato numero di stati.

I dati della tabella sono in generale significativi ed affidabili fatta eccezione per le decomposizioni elencate qui di seguito in cui non è possibile distinguere in modo preciso gli stati e quindi non offrono né un bon livello di dettaglio né una stima precisa. Le eccezioni sono le seguenti:

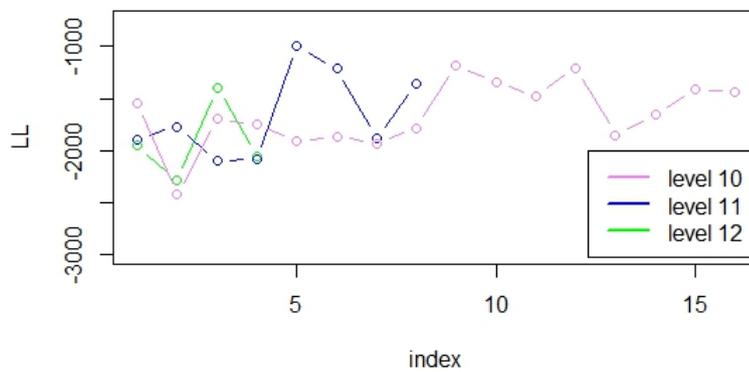
A. DATI RIASSUNTIVI DEI MODELLI A CONFRONTO

- 3 stati
 - level 10, index 0
 - level 10, index 1

- 4 stati
 - level 11, index 0
 - level 11, index 4
 - level 10, index 0
 - level 10, index 1
 - level 10, index 8
 - level 10, index 9



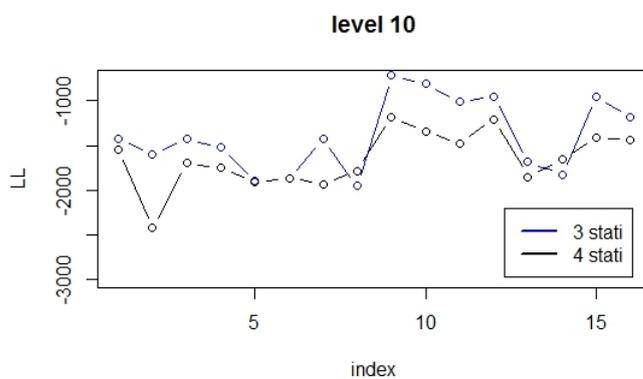
(a) 3 stati



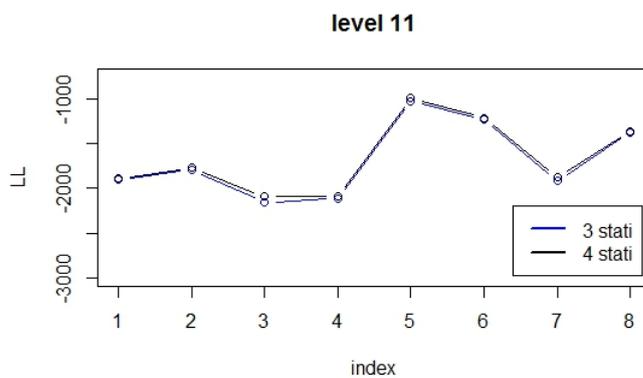
(b) 4 stati

Figura A.1: Rappresentazione grafica dei coefficienti di LL rispetto al tipo di modello (a) 3 stati o (b) 4 stati.

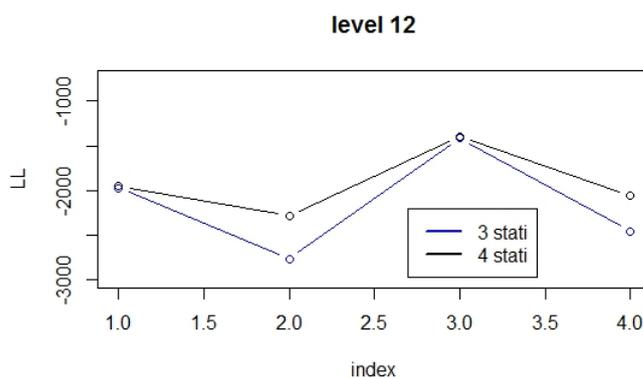
A. DATI RIASSUNTIVI DEI MODELLI A CONFRONTO



(a) level 10



(b) level 11



(c) level 12

Figura A.2: Rappresentazione grafica dei coefficienti di LL rispetto al ciascun livello separatamente.

Ringraziamenti

Desidero ringraziare il professor Matteo Golfarelli per il tempo e la pazienza dedicatami e per l'interesse che ha saputo trasmettermi per questa materia.

Voglio ringraziare i nuovi amici e compagni trovati durante questo percorso per gli indimenticabili momenti passati insieme in questa parte del mio cammino universitario, sarebbe stato per me molto più difficile e noioso se non avessi avuto il vostro aiuto e la vostra amicizia.

Per ultimi ma non per importanza ringrazio la mia famiglia, Lorenzo, Giada e tutti gli amici di sempre su cui so di poter contare.

Bibliografia

- [1] Daniel T.L. Lee & Akio Yamamoto, "*Wavelets Analysis: Theory and Applications*", Hewlett-Packard Journal, December 1994.

- [2] M. Y. Gokhale & Daljeet Kaur Khanduja, "*Time Domain Signal Analysis Using Wavelet Packet Decomposition Approach*", Int. J. Communications, Network and System Sciences, 2010, 3, 321–329.

- [3] Rachel E. Learned & Alan S. Willsky "*A Wavelets Packet Approach to Transient Signal Classification*", Applied and computational harmonic analysis 2, 265-278 (1995).

- [4] Oliver Rioul & Martin Vetterli, "*Wavelets and Signal Processing*", IEEE SP Magazine, October 1991.

- [5] Diego Tobon-Mejia, Kamal Medjaher, Nouredine Zerhouni, Gerard Tripot, "*A data-driven failure prognostics method based on mixture of gaussian hidden markov models*", HAL, October 2012.

- [6] Jafar Zarei & Javad Poshtan, "*Bearing fault detection using wavelets packet transform of induction motor stator current*", Tribology International 40 (2007) 763-769.

- [7] Hai Qiu, Jay Lee, Jing Lin, Gang Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics", *Journal of Sound and Vibration* 289 (2006) 1066-1090.
- [8] Ming Dong & David He, "A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology", *Mechanical Systems and Signal Processing* 21 (2007) 2248-2266.
- [9] L.R. Rabiner & B.H. Juang, "An Introduction to Hidden Markov Models", *IEE ASSP MAGAZINE*, gennaio 1986.
- [10] G.L.Kouemou, "History and Theoretical Basics of Hidden Markov Models", ISBN: 978-953-307-208-1, InTech (2011).
- [11] Liao, Lee, "Design of a Reconfigurable Prognostics platform for Machine Tools", *Expert Systems with Applications* (2009) 37: 240–252.
- [12] Gray Do., Rivers D., Vermont G., "Measuring the Economic Impacts of NSF I/UCRC Program: A Feasibility Study", (2012).
- [13] Wang H.; Lee J.; Ueda T.; Adjallah H.H.; Ghaffari M. "Engine Health Assessment and Prediction Using the Group Method of Data Handling and the Method of Match Matrix – Autoregressive Moving Average", *Proceeding of 2007 ASME Turbo Expo*.
- [14] Lawrence R. Rabiner, *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, IEEE, vol.77, no.2, 1989.
- [15] L. Dielle, C. Berenguer, A. Frall e M. Roussignol *Sequential condition-based maintenance scheduling for a deteriorating system*

BIBLIOGRAFIA

- [16] Mike Barlow, *Predictive Maintenance, A World of Zero Unplanned Downtime*, O'REILLY, 2015.
- [17] Brian W. *Intelligent building care, Facilities* , Volume 17, Number 5/6 , pp. 189–194 (1999).
- [18] Moubray *Reliability-Centered Maintenance* (2001).
- [19] AFNOR, “*Condition monitoring and diagnostics of machines prognostics - part 1: General guidelines*. NF ISO 13381-1,” 2005.
- [20] <https://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/publications/>
- [21] www.jethroproject.com, TJP, November 10, 2008, pp. 1-5 *Condition Based Maintenance*, Copyright 2008 TJP, Byron A. Ellis
- [22] Jianbo Yu, “*Health Condition Monitoring of Machines Based on Hidden Markov Model and Contriution Analysis*”, IEEE, 14 February 2012.
- [23] Jianbo Yu, “*Local and Non local Preserving Projection for Bearing Defect Classification and Performance Assessment*”, IEEE, 03 October 2011.
- [24] Eker, Ömer, Faruk, Camci, Faith, Jennions, Ian K., “*Major Challenges in Prognostics: Study on Banchmarking Prognostics Dataset*”, CERES, 31 December 2012.
- [25] Sergey Porotsky, Zigmund Bluvband ALD Group, “*Remaining Useful Life Estimation for System with Non-Trendability Behaviour*”, IEEE, 13 September 2012.
- [26] Soumaya Yacout, “*Logical Analysis of Maintenance and Performance Data of Physical Assets*”, IEEE, 03 April 2012.

- [27] David Martínez-Rego, Oscar Fontenla-Romero, Amparo Alonso-Betanzos, "*Power Wind Mill Fault Detection via one-class ν -SVM Vibration Signal Analysis*", IEEE, 03 October 2011.
- [28] Mohamad-Ali Mortada, Soumaya Yacout, "*ebmLAD-using Logical Analysis of Data in Condition Based Maintenance*", IEEE, 05 May 2011.
- [29] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, "*Hidden Markov models for failure diagnostics and prognostic*", IEEE, 05 July 2011.
- [30] Fengtao Wang, Yangyang Zhang, Bin Zhang, "*Application of Wavelet Packet Sample Entropy in the Forecast of Rolling Element Bearing Fault Trend*", IEEE, 18 July 2011.
- [31] Diego A. Tobon-Mejia, Kamal Medjaher, Noureddine Zerhouni, "*A mixture of gaussian hidden markov model for failure diagnostic and prognostic*", IEEE, 21 October 2010.
- [32] H. Qiu, J. Lee, J. Lin, G. Yu, "*Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics*", Journal of sound and vibration, 03 October 2011.
- [33] M. Dong and D. He, "*A segmental hidden semi-markov model (hsmm)-based diagnostics and prognostics framework and methodology*", Mechanical Systems and Signal Processing, vol. 21, pp. 2248–2266, 2007.